



PHD

Spatio-Temporal modelling of Dengue Fever in Zulia state, Venezuela

Cabrera, Maritza

Award date:
2013

Awarding institution:
University of Bath

[Link to publication](#)

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

Take down policy

If you consider content within Bath's Research Portal to be in breach of UK law, please contact: openaccess@bath.ac.uk with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

Spatio-Temporal modelling of Dengue Fever in Zulia state, Venezuela

submitted by

Maritza Cabrera

for the degree of Doctor of Philosophy

of the

University of Bath

Department for Health

December 2013

COPYRIGHT

Attention is drawn to the fact that copyright of this thesis rests with its author. This copy of the thesis has been supplied on the condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without the prior written consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purposes of consultation.

Signature of Author

Maritza Cabrera

Abstract

Over half of the world's population are at risk of infection from dengue fever (Guha-Sapir and Schimmer, 2005). This viral disease is transmitted by the female *Aedes aegypti* mosquito and is the major source of human death in the world when compared with any other vector borne disease (Gubler, 1998a). The first important epidemic of dengue haemorrhagic fever (DHF) in America was reported in Cuba in 1981 and subsequently in Venezuela during 1989 and 1990 (Oletta, 2006; Brightmer and Fantato, 1998). There has been a trend of increased incidence in many Central and South American countries since 1990 - Brazil, Venezuela, Honduras and Mexico (SanMartin et al., 2010) with Venezuela having the highest number of cases of DHF. The urgent need for more effective public health measures to combat this disease in Venezuela drove the decision to undertake the work described in this dissertation.

Spatio-Temporal modelling has been developed for the prediction of the occurrence of dengue fever in Zulia state, Venezuela. A systematic approach has been adopted to validate this tool. At the first stage of the analysis an exploratory study was performed to underline the most significant features of the dynamics of incidence rates of dengue fever from 2002 to 2008. In the second stage a Generalized Linear Model (GLM) approach was used in the form of Negative Binomial Generalized Linear Mixed model (GLMM) to compare Relative Risk (RR) across exposure groups by age and sex, using an epidemiological dataset covering the whole of Zulia State, Venezuela. This approach used both a frequentist and a Bayesian perspective for comparative purposes of both outcomes and methodologies. Finally a Spatio-Temporal model was constructed based

on Generalized Additive Mixed model (GAMM) framework because the earlier analysis identified a complex association between covariates and response variables. This GAMM structure was further developed so that it could be used to help predict future outbreaks of the disease in Zulia state with a good degree of accuracy.

Key words: dengue fever, Spatio-Temporal modelling, Generalized Linear Mixed model (GLMM), Generalized Additive Mixed model (GAMM).

Acknowledgements

This thesis has represented another marvellous chapter in my life and is now complete. A new chapter is opening up. No doubt it will be full of new challenges and adventures. The thesis could not have been written without the companionship of many colleagues and friends who have accompanied every single page of it.

Firstly, I want to thank the Department for Health at the University of Bath for giving me the opportunity to make my academic dream a reality. For the department of Mathematics for trusting me as a tutor in a variety of subjects in the field of probability and statistics, and for the Department of Economics for adopting me as a tutor for an extraordinary group of students. I offer my sincere thanks to all those who contributed to this project from its earliest conception. This includes Ministry of Health for the Zulia state of Venezuela which provided me with invaluable epidemiological data, especially to Dr Belen Rodriguez for her unconditional support. The Ministry for the Environment of the Zulia State gave me the available information I needed to conduct my research. I would particularly like to acknowledge the encouragement of Dr Alejandro Gonzalez for his friendship. The initial conversations I had with Dr Lelys Bravo from the Simon Bolivar University and Dr Yazmin Rubio from the University of Carabobo proved to be very fruitful for the maturation of the argument. In Bath, I have benefited in the Department of Mathematics from the support of Dr Gavin Shaddick who believed in my project and encouraged me to bring it to fruition; especially to Dr Merrilee Hurn who encouraged me during the most difficult times of my journey and gave me the impetus to carry it forward. My gratitude and friendship will be

ongoing. I was fortunate to cross the path of Professor Peter Diggle whose passion for epidemiology made statistics alive and responsive to the real needs of people. His inspiration has encouraged me to adopt the same commitment through research and teaching. Last but not least, this thesis would never have been submitted without the support of my supervisors: Dr Michelle Sims who made time to comment on my work and Dr Gordon Taylor, who allowed me to pursue freely my ideas giving me advice whenever it was needed.

But writing a PhD is not only an academic endeavour, it is also about spiritual and emotional growth. This is why this doctoral journey would never have been possible without friends and family to support it at every stage.

I was fortunate to have been offered an individual office at the University and forged many friendships with people who popped into my office for a chat and a cup of tea. It would be impossible to name them all but I want to highlight the colleagues and friends who have shared their enthusiasm, experiences, joys and sorrows: Alan, Anna, Cathy Fl., Cathy Fr., David P., Elaine and David, Elisa, Fiona, Jason, Lisa, Maxime, Severine and Silvy. My special gratitude to Rebecca Bruce and Susanna Jarvis from the International Office for her outstanding support and friendship. The community of St Johns and its choir gave me an extraordinary support network and opened for me the door to the English culture: Ann, Amina, Angelina, Catherine, David, Joy, Julia, Malcolm, Neil, Peter, Rupert, Ruth and Richard, Sheila, Sue, Tina and Johannes, Vanessa, Ula and Anne Louise who unfortunately passed away during the PhD process. With them we shared musical productions with the BBC and the recording of a CD, and many laughs and joyful moments. In the City of Bath, I encountered many people who became friends and a great source of support: Ana Maria, Betty, Christine and Peter with their family, Irene M. and Mario with their girls, Irene P. and her family, Julie, Karen and her kids, Max and Anne, Pascale, Zdenka, and many others to whom I am indebted. Jason and Christine reviewed the entire manuscript and corrected many English spelling and grammatical mistakes. They were always cheerful, positive and

constructive despite the hard work. Thanks particularly to Dr David Packham because of the final reading of my dissertation and his wise counsels.

In Venezuela, it was the scholarship that Fonacit-Fundayacucho granted which enabled me to come to Bath in the first place and enrol into the PhD programme.

My family offered me the "carino" every day that any Latino in England needs in order to be myself in a foreign environment. I thank my parents for their invaluable support, generosity, love and friendship. They have been my rock and my greatest inspiration. Every member of my extended family has a special place in this PhD project. But special tribute to my nieces and nephews: Angel, Cesar, Claudia, Diego, Laura, Marcos y Yuliana. One day, they will spread their wings and fly towards freedom of thought following their own dreams.

Finally I thank the One who makes me stand up in times of adversity, so my ankles never were weakened (Psalm 18).

Contents

1	Introduction	18
1.1	Infectious disease of dengue fever	18
1.2	The health impact of dengue fever in Venezuela	20
1.3	The problem	22
1.4	Aims	23
1.5	Research question	23
1.6	Significance of the project	23
1.7	Thesis Overview	25
2	Theoretical background	28
2.1	Introduction	28
2.2	Literature review	29
2.3	Surveillance and control of infectious diseases	31
2.3.1	Quality and validity of spatial data	34
2.3.2	Georeferenced data of dengue fever in Zulia state	34
2.4	Statistical methods for epidemiological studies	43
2.4.1	Generalized Linear model (GLM) for count data	45
2.4.2	Generalized Linear Mixed models (GLMM) for count data	48
2.4.3	Disease mapping	51
2.4.4	Generalized Additive models (GAM) for count data	54
2.5	Explanatory factors associated with dengue fever	59
2.5.1	Individual factors	59

2.5.2	Specific serotype circulating	60
2.5.3	Climatic factors	60
2.5.4	Socioeconomic factors	69
2.5.5	Geographic factors	72
2.5.6	Demographic factors	72
2.6	Framework of Early Warning System (EWS) for infectious diseases . . .	73
2.6.1	Protocol for developing Early Warning Systems for infectious dis- eases	74
2.6.2	A review of Early Warning System of Dengue fever	76
3	Exploratory data analysis	79
3.1	Introduction	79
3.2	Research objectives	80
3.3	Research questions	80
3.4	Data exploration of the dynamics of dengue fever in Zulia-state	81
3.4.1	Data exploration using individual factors	87
3.4.2	Data exploration with geographic factors	91
3.4.3	Data exploration using socioeconomic factors	102
3.4.4	Data exploration using demographic factors	105
3.5	Data exploration using climatic factors	108
3.5.1	Data exploration using temperature indexes	108
3.5.2	Data exploration using rainfall indexes	112
3.5.3	Data exploration using the El Nino3.4 index	115
3.6	Discussion and conclusions	116
4	GLM: age and gender effects of dengue fever in Zulia state	120
4.1	Introduction	120
4.2	Research objective	122
4.3	Research question	122
4.4	Model specification via a frequentist approach	122
4.5	Analysis of results from the frequentist approach	124

4.6	Model specification via a Hierarchical Bayesian approach	126
4.7	Analysis of results from the Bayesian approach	129
4.8	Comparing limitations and benefits from both methodologies	135
4.9	Discussion and conclusions	136
5	GAM: Spatio-temporal modelling of dengue fever, towards an EWS	140
5.1	Introduction	140
5.2	Research objectives	142
5.3	Research questions	142
5.4	Model specification	143
5.5	Analysis of results	149
5.6	Predicting dengue cases in Zulia state	162
5.7	Discussion and conclusions	166
6	General discussion and conclusion	170
6.1	Introduction	170
6.2	Discussion and conclusions	171
6.3	Limitations	176
6.4	Future directions	177
A	Exploratory analyses	180
A.1	Exploratory analysis of socioeconomic factors	180
A.2	Multipanel of socioeconomic factors aggregated at municipality level . .	184
A.3	Multipanel scatterplots of socioeconomic factors in the whole of Zulia .	189
A.4	Scatterplot showing the association between temperature indexes	190
B	Model specifications	195
B.1	Model specification via a frequentist approach	195
B.2	Model specification via a Hierarchical Bayesian approach	196
B.3	Analysis of results from the Bayesian approach	199

C	GAM: Spatio-Temporal modelling	201
C.1	First stage: GAMs for lagged effects	201
C.2	Second stage: GAMM with municipality as a random effect	209
C.3	Third stage: GAMM for predictive purposes	228

List of Figures

1.1	Geographic representation of Venezuela with state boundaries	21
1.2	Overview of the thesis	27
2.1	Search strategy on the literature review of this project	29
2.2	Zulia state with its municipalities boundaries: (1) Padilla, (2) Paez, (3) Mara, (4) Lossada, (5) Maracaibo, (6) San Francisco, (7) Urdaneta, (8) Santa Rita, (9) Cabimas, (10) Simon Bolivar, (11) Lagunillas, (12) Valmore Rodriguez, (13) Baralt, (14) Miranda, (15) Rosario, (16) Machiques, (17) Semprun, (18) Catatumbo, (19) Colon, (20) Pulgar, (21) Sucre . .	36
2.3	Geographical regions of the El Nino Southern Oscillation (ENSO) on the Pacific Ocean	63
2.4	Protocol for developing Early Warning System (EWS) of infectious disease	75
3.1	The five main administrative regions which comprise Zulia state	82
3.2	Distribution of dengue incidence weekly aggregated per 10,000 inhabitants in Zulia from 2002-2008	84
3.3	Mapping of crude incidence rates of dengue per 10,000 in Zulia state from 2002 to 2008	86
3.4	Observed incidence rate of dengue fever weekly aggregated by age group in Zulia from 2002 to 2008	88
3.5	Observed incidence rates of dengue fever weekly aggregated by gender in the whole of Zulia from 2002 to 2008	90

3.6	Evolution of incidence rate in the Guajira region, monthly aggregated from 2002 to 2008	92
3.7	Evolution of incidence monthly aggregated in East Coast region from 2002 to 2008	93
3.8	Evolution of incidence monthly aggregated in Capital region from 2002 to 2008	94
3.9	Evolution of incidence monthly aggregated in Perija region from 2002 to 2008	95
3.10	Evolution of incidence monthly aggregated in South region from 2002 to 2008	96
3.11	Spatial distribution of average altitude across municipalities	98
3.12	Scatterplot of Average altitude versus incidence rates of dengue fever at municipality level	99
3.13	Geographical locations of municipalities with territory coverage, expressed in km^2 : (1) Rosario, (2) Pulgar, (3) Semprum, (4) Colon, (5) Cabimas, (6) Valmore, (7) Urdaneta, (8) Padilla, (9) Lossada, (10) Baralt, (11) Mara and (12) Catatumbo in Zulia state	100
3.14	Geographical locations of municipalities with territory coverage, expressed in km^2 : (13) Lagunillas, (14) Paez, (15) Maracaibo, (16) Miranda, (17) Machiques, (18) Francisco, (19) Sta. Rita, (20) Simon Bolivar and (21) Sucre in Zulia state	101
3.15	Spatial distribution on percentage of households with access to the pipe water supply network	103
3.16	Scatterplot of percentage of households with access to water pipe supply network versus the incidence rates of dengue fever aggregated at municipality level	104
3.17	Time series of monthly aggregated average temperature per grids in Zulia state	109
3.18	Time series of monthly aggregated maximum temperature per grids in Zulia state	110

3.19	Scatterplot of monthly aggregated maximum temperature averaged per grid versus dengue cases in the whole of Zulia	112
3.20	Time series of monthly aggregated average rainfall per grids in Zulia state	113
3.21	Scatterplot of monthly aggregated average rainfall averaged per grid versus dengue cases in the whole of Zulia	114
3.22	Scatterplot of monthly aggregated the El Nino3.4 (6 months lag delay) versus dengue cases in the whole of Zulia	116
4.1	Confidence intervals (95%) of RR by age groups and years	125
4.2	WinBUGS outputs showing the Posterior kernel densities of b[1] year 2003, b[2] year 2004 and b[7] group of age from 5 to 14 years old.	130
4.3	WinBUGS outputs showing the Dynamic traces of b[1] year 2003, b[2] year 2004 and b[7] group of age from 5 to 14 years old.	131
5.1	GAM-smooth curves with 95% confidence intervals from the optimal structure at the first stage	152
5.2	GAMM estimated functions using random effect at municipality level, with 95% confidence interval from the second stage	155
5.3	Mapping Moran's I: Random spatial autocorrelation from 2002 to 2008 from GAMM structure at the second stage	157
5.4	Estimated smooth curves from the final GAMM structure for predictive purposes	160
5.5	Predictive trend line from the final GAMM with 90% predictive interval using data from July 2009 to June 2010.	164
6.1	Directions for future work	178
A.1	Scatterplot of percentage of households which have a garbage collection service against incidence of dengue fever	180
A.2	Scatterplot of households which are unoccupied against incidence rate of dengue	181

A.3	Scatterplot of percentage of households with residential air conditioning equipment against incidence rate of dengue	182
A.4	Scatterplot of percentage of urban population against incidence of dengue fever	183
A.5	Multipanel of socioeconomic variables in Valmore: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment	184
A.6	Multipanel of socioeconomic variables in Baralt: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment	185
A.7	Multipanel of socioeconomic variables in Maracaibo: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment	186
A.8	Multipanel of socioeconomic variables in Cabimas: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment	187

A.9	Multipanel of socioeconomic variables in Lagunillas: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment	188
A.10	Multipanel scatterplots of socioeconomic factors with Pearson's coefficients amongst them: (1) No. of households with access to the water supply network, (2) No. of homes which have garbage collection service, (3) No. of households with access to electricity, (4) No. of households which are unoccupied, (5) No. of households including a bathroom without basic shower, (6) No. of households without basic living standards .	189
A.11	Multipanel plots of temperature indexes with 1 month time lag in Baralt	190
A.12	Multipanel plots of temperature indexes with 1 month time lag in Valmore	191
A.13	Multipanel plots of temperature indexes with 1 month time lag in Maracaibo	192
A.14	Multipanel plots of temperature indexes with 1 month time lag in Cabi-mas	193
A.15	Multipanel plots of temperature indexes with 1 month time lag in Lagunillas	194
B.1	WinBUGS outputs of dengue incidence in Zulia state. Autocorrelation functions	199
B.2	Bivariate posterior scatter plots	200
C.1	Pearson and Deviance residual against fitted values from the optimal model with free restriction on the degree of freedom	204
C.2	Deviance residual split up at regional level	205
C.3	Deviance residual against mean rainfall from the optimal model with free restriction on the degree of freedom	206

C.4	Deviance residual against the El Nino3.4 from the optimal model with free restriction on the degree of freedom	207
C.5	Deviance residual against maximun temperature from the optimal model with free restriction on the degree of freedom	208
C.6	ACF from optimal model with free restriction on the degree of freedom	209
C.7	Deviance of GAMM with random effect - Maracaibo	210
C.8	Deviance of GAMM with random effect - Cabimas	211
C.9	Deviance of GAMM with random effect - Lagunillas	211
C.10	Deviance of GAMM with random effect - Colon	212
C.11	Deviance of GAMM with random effect - Catatumbo	212
C.12	Deviance of GAMM with random effect - Urdaneta	213
C.13	Deviance of GAMM with random effect - Francisco	213
C.14	Deviance of GAMM with random effect - Mara	214
C.15	Deviance of GAMM with random effect - Miranda	214
C.16	Deviance of GAMM with random effect - Rosario	215
C.17	Deviance of GAMM with random effect - Baralt	215
C.18	Deviance of GAMM with random effect - Sucre	216
C.19	Deviance of GAMM with random effect - Pulgar	216
C.20	Deviance of GAMM with random effect - Machiques	217
C.21	Deviance of GAMM with random effect - Padilla	217
C.22	Deviance of GAMM with random effect - Lossada	218
C.23	Deviance of GAMM with random effect - Semprum	218
C.24	Deviance of GAMM with random effect - Rita	219
C.25	Deviance of GAMM with random effect - Rodriguez	219

List of Tables

1.1	Clinical stages of dengue infection.	19
2.1	Geographical location of municipalities by Grid I, II, III and IV	41
2.2	Initial data set georeferenced in Zulia state provided by local agencies	42
2.3	Protocol for data exploration including techniques before fitted models	44
2.4	Overview of parametric Poisson GLM for count data	46
2.5	Temporal evolution of dengue transmission in human population	65
2.6	Climatic predictors with appropriate time lag associated with dengue fever	66
2.7	Other aetiological factors associated with dengue fever in a socioeconom- ical context	70
3.1	Estimated population annually aggregated per municipality from 2002 to 2008	106
3.2	Pearson's coefficients of potential non-climatic factors associated with the incidence rate of dengue fever: (1) Percentage of unoccupied house- holds, (2) Percentage of households that receive garbage collection ser- vice, (3) Percentage of households with access to the pipe water supply network, (4) Altitude, (5) Population density, (6) Urban population, (7) Percentage of households that receive electricity service, (8) Percentage of households with domestic air conditioning equipment	108
3.3	Maximum temperatures settled by GRID I, II, III, IV	111
3.4	Average of minimum temperature settled by GRID I, II, III, IV	112
4.1	Negative Binomial estimates via a frequentist approach	124

4.2	Posterior summary of the Bayesian Hierarchical Poisson approach using WinGUBS	133
4.3	Comparison between both the Frequentist NB approach and the Bayesian NB approach	134
5.1	Climatic covariates with lagged effects to be modelled	144
5.2	Outcomes of the GAM structure at the first stage: optimal lagged effects	151
5.3	Outcomes of the GAMM using municipality as a random effect from the second stage	154
5.4	Outcomes from the final GAMM structure for predictive purposes . . .	159
5.5	Binomial predictive interval based on the final GAMM structure, at 90% CI, using data from July 2009 to June 2010	165

Chapter 1

Introduction

1.1 Infectious disease of dengue fever

As with many infectious diseases in the human population, dengue fever is a vector borne disease, in which the virus is transmitted by the bite of a female mosquito belonging to the *Aedes aegypti* species. More than 2.5 billion people (over half of the world's population) are at risk of being infected with dengue fever (Guha-Sapir and Schimmer, 2005). Scientific reports have stated that it is the major source of human death in the world when compared with any other vector borne disease (Gubler, 1998a). The virus of dengue fever can be transmitted in both directions: mosquito to human and vice versa. The mosquito, once infected remains so for the rest of her life (Gibbons and Vaughn, 2002). There are four serotypes of the virus named DEN-1, DEN-2, DEN-3 and DEN-4 and a person who has suffered with one of them does not get immunity from the other three. Furthermore, any of these serotypes can develop into dengue haemorrhagic fever (DHF), considered the most dangerous stage of the disease. This haemorrhagic stage is characterized by circulatory collapse, leading to death (Gibbons and Vaughn, 2002; Raja and Devi, 2006) in most of the cases. Regarding the clinical stages of the disease, they can be classified into: classical dengue fever (DF), dengue haemorrhagic fever (DHF) and dengue shock syndrome (DSS). The symptoms (Gibbons and Vaughn, 2002; Soni et al., 2001; Gubler, 1998a) of which, are summarized in Table 1.1.

Dengue virus	Pathology
DF	Headache, red skin eruption, muscular fatigue, reduction in white blood cells.
DHF	Evident haemorrhagic manifestations through mucous, injection site or gastrointestinal zone. Reduction in platelet counts less than 100,000 compared to normal range of 150,000 to 440,000 per cubic millimetre.
DSS	Circulatory collapse, causing deaths from 8 to 24 hours after onset. Pulse pressure descend drastically to 20 mm Hg compared to a normal value of 40.

Table 1.1: Clinical stages of dengue infection.

Epidemiological records show that dengue fever has caused a large number of deaths in the human population across the world since the 17th century (Gubler, 1998b). However according to these authors, it was not until the 1960's that control programs to prevent the breeding of the mosquito responsible for the disease were implemented worldwide through the use of chemical insecticides. This situation remained relatively stable for a period of time but by the late 1990s a resurgence of the disease caused a global alarm in the public health sector (Gubler, 1998a). Moreover the gravity of this issue became worse by the fact than only nine countries reported cases in the 1950's, whereas today more than 100 countries have had experience of the disease (Guha-Sapir and Schimmer, 2005; SanMartin et al., 2010), covering: Asia, the Pacific, the Americas, Africa, the Middle East and the Caribbean.

In addition although dengue fever has been endemic in most tropical and subtropical countries (Gibbons and Vaughn, 2002), more recently there has been a large number of cases in virgin territory (Gibbons and Vaughn, 2002; Raja and Devi, 2006; Gurtler et al., 2009). This pattern has been led due to the migration of the mosquito species searching for a more favourable environment for feeding and breeding (Wu et al., 2009; Lowe et al., 2011).

However the aetiology of the resurgence of the disease is still complex to understand but it might involve: different serotypes circulating, lack of effectiveness of mosquito control programmes, unplanned urbanization, ineffective public health infrastructures, sociodemographic diversity and climate change amongst others (Gubler, 1998a; Gibbons and Vaughn, 2002; SanMartin et al., 2010; Bohra and Andrianasolo, 2001; Lowe et al., 2011).

1.2 The health impact of dengue fever in Venezuela

An increase of dengue infection observed in Latin America in the 1970s was owing to the lack of control programmes to eradicate the breeding sites of the mosquito population (Gubler, 1998b). The first important epidemic of dengue haemorrhagic fever (DHF) in America was reported in Cuba in 1981 and the second affected Venezuela during 1989 and 1990 (Oletta, 2006; Brightmer and Fantato, 1998). Furthermore a study of the Americas revealed a general trend of increased dengue incidence in: Brazil, Venezuela, Honduras and Mexico, during the period 1990 to 2007 (SanMartin et al., 2010). According to the authors, Venezuela reported the highest number of cases of dengue haemorrhagic fever (DHF).

This fact demonstrates that Venezuela is a country at serious risk regarding the spread of the disease. The study also revealed that the severity of the disease in Venezuela reached a peak in 2007, (SanMartin et al., 2010). Thus suggesting that some of the aforementioned causes are linked in some way. In addition, an outbreak of the disease in 2006 was particularly severe in Zulia state located in the north west of the country (Oletta, 2006) (see Figure 1.1). Furthermore, it was reported by the National Ministry of Health that Zulia state experienced the highest number of cases nationwide during 2007 (25% of all the cases), followed by Aragua located in the north central area with 7% of the cases.



Figure 1.1: Geographic representation of Venezuela with state boundaries

The main concern is that even though the spread of dengue has been researched in a number of areas within Venezuela, including Aragua (Barrera et al., 2002; Rodriguez et al., 2007) and the Capital District (Saez, 2006), both located on the central part of the country, a paucity of research is still observed covering Zulia state (Anez et al., 2006).

1.3 The problem

Taking Venezuela as a whole, some authors have highlighted a paucity of studies on dengue fever given the current alarming situation seen in the country (Brightmer and Fantato, 1998). In particular, scientific evidence needs to be revealed at spatial and temporal scale regarding dengue fever across different regions of Venezuela. Special attention merits Zulia, which has been considered by some experts as an endemic region for the past 15 years (Anez et al., 2006; Oletta, 2006). But even more, the trends remain critical during 2009 and 2010 (Mojica et al., 2012).

From an economic perspective, a study carried out in Zulia state during 1997 to 2003 revealed that 64% of the financial deficit in the state could be attributed to labour absenteeism owing to reported cases during this period (Anez et al., 2006). These authors stated that the loss to the economy, totalling US\$ 1,348,077 may be equivalent of 20 thousand new jobs annually offered. In summary, the public health and socio-economic issues show the urgent need for scientific initiatives in this matter.

Hence the goal of this research is to develop an extensive epidemiological study of dengue fever in Zulia state at space and temporal scales with the main purpose of predicting epidemics in advance. As a result, a valuable tool shall be provided to the health authorities in Venezuela that could be used to target effective programmes for eradicating dengue fever in the north west part of Venezuela.

The recent rise of spatial epidemiological studies focussed on dengue fever worldwide (Lowe et al., 2011; Muttitanon et al., 2004; Mammen et al., 2008; Hales et al., 1999), and some local initiatives conducted in Venezuela (Saez, 2006; Rodriguez et al., 2007; Barrera et al., 2002; Rifakis et al., 2005; Barrera et al., 2000; Monsalve et al., 2010) allow the discussion of advantages and disadvantages of methodologies applied, seeking for a better understanding on the spread of the disease using the most appropriate technique.

Although the use of spatial clustering of the disease following the disease mapping technique has become a standard procedure (Lawson et al., 2000; Lawson and Fiona, 2001; Lawson, 2009) to identify areas at risk. New approaches are emerging as potential tools (Wood, 2006; Zuur, 2012) in the field of Spatial Epidemiology to explain the complexity on the transmission of the diseases.

1.4 Aims

The aims of the present project are:

- To develop a Spatio-Temporal modelling of dengue fever in Zulia state from 2002 to 2008 via a generalized additive model with random effects.
- Use the previous model for predictive purposes towards the construction of an Early Warning System (EWS) of dengue fever in Zulia state.

1.5 Research question

- What is the most effective way of Spatio-Temporal modelling of dengue fever with a flexible approach to capture the real dynamic associated with a set of climatic and non-climatic covariates?
- Can a predictive model of dengue fever be implemented in Zulia state, Venezuela where climatic data is limited?

1.6 Significance of the project

Scientific evidence places Venezuela in a high risk level of dengue transmission (Brightmer and Fantato, 1998; Anez et al., 2006; Oletta, 2006; Barrera et al., 2000). And a serious public health issue remains over time (Mojica et al., 2012) without an integrated initiative in the fight against dengue fever. Academics in the field of epidemiology have

emphasized the potential role of statistical methods in the public health sector, for effectively targeting control and prevention programs of the diseases (Chang et al., 2009; Chan, 2012).

Although epidemiological studies related to dengue fever concerning Venezuela has been comparatively sparse (Brightmer and Fantato, 1998), some studies have been carried out in order to understand the aetiology of the disease. Methodologies based on multiple regression and Pearsons coefficients have assessed the association between climatic data and health outcome of dengue cases (Saez, 2006; Rodriguez et al., 2007; Barrera et al., 2002). However only one study has focused on the influence of the El Nino Southern Oscillation (ENSO) using epidemiological data from Venezuela (Rifakis et al., 2005). This study revealed no- evidence of association, perhaps because delay time was not considered or because the statistical approach implemented was not the most appropriate.

Some other studies based on multiple regression have been undertaken to identify factors associated with the spread of the disease in Venezuela (Barrera et al., 2000, 2002). These findings revealed that: human population density, lack of water supply and numbers of *Aedes aegypti* females resting inside households are statistically significant. Finally a study using a Bayesian spatio-temporal model of dengue fever in Venezuela was conducted in Maracay, Venezuela (Monsalve et al., 2010). However some technical concerns regarding spatial and temporal autocorrelations (Pfeiffer et al., 2008) need to be included into the analysis in order to justify such implementation. Furthermore considerations regarding delay time for climatic covariates were not considered in the study performed in Maracay.

As a result the current project brings together for the first time different sources of information available to pursue a spatio-temporal modelling of dengue fever in Zulia state, Venezuela. This project has also greatly benefited from remotely sensed climatic data provided by the international agency of the National Aeronautics and

Space Administration's (NASA) (Kempler, 2013). To this end an exploratory analysis was conducted to investigate patterns of the dynamic of dengue fever in Zulia state in conjunction with potential factors associated with the disease. At this stage collinearity amongst the explanatory variables was also assessed to assist the construction of more advanced techniques.

Subsequently a Generalized Linear model for the count dengue cases was designed to compare Relative Risk (RR) by exposure groups of gender and age via a frequentist and Bayesian perspectives for comparative purposes. And finally a spatio-temporal modelling of dengue fever was conducted via a Generalized Additive structure to capture the complexity of the associations regarding the climatic covariates. At this stage a set of non-climatic covariates and climatic factors of temperature, rainfall and an index of the El Nino Southern Oscillation (ENSO) with appropriated lagged effects for all of them were included within the model. Furthermore the Generalized Additive structure was extended into a Generalized Additive Mixed model (GAMM) using municipality as a random effect. However tests assessing spatial and temporal autocorrelations were previously evaluated. Consequently this structure was used for predictive purposes of dengue cases in Zulia state validated with a high standard of accuracy to be extended into an Early Warning System for further studies.

1.7 Thesis Overview

The present thesis, summarized in Figure 1.2 comprises six chapters. Chapter 1, discusses the overall problem along with general aims and questions to be elucidated. Chapter 2, shows the theoretical background on the field of spatial epidemiology revealing openness to the available methodologies and most relevant results aimed to justify the statistical approaches followed in the present investigation. Chapters 3, 4 and 5 each have their own specific objectives and although independent they complement each other in achieving the overall goal.

Chapter 3, presents in great detail an exploratory analysis of dengue incidence in Zulia state, providing evidences of implementing more robust and advanced statistical techniques according to the suggestion driven by the data. The subsequent approach discussed in Chapter 4, was based on estimating the relative risk of dengue fever by comparing different exposure groups at age and gender levels with data aggregated in the whole Zulia. A classical frequentist analysis and a hierarchical Bayesian approach were implemented using generalized linear framework for count data for comparative purposes. Chapter 5, adopts a semiparametric approach of the incidence of dengue fever in Zulia state using a novel statistical technique of generalized additive model showing the complexity associated with climatic and non- climatic covariates and the effect on incidence of the disease. Toward the construction of an Early Warning System (EWS) in Zulia state, Venezuela, previous findings of the GAM were used for modelling predictive cases of dengue with a high standard of accuracy. The final model was based not only on climatic but also on non- climatic covariates for a more robust understanding of the dynamics of dengue transmission and was validated with high standard of accuracy.

Finally, chapter 6 puts forward a global conclusion and discussion based on the overall goals submitted in Chapter 1, by considering the conclusions and findings discussed at every chapter.

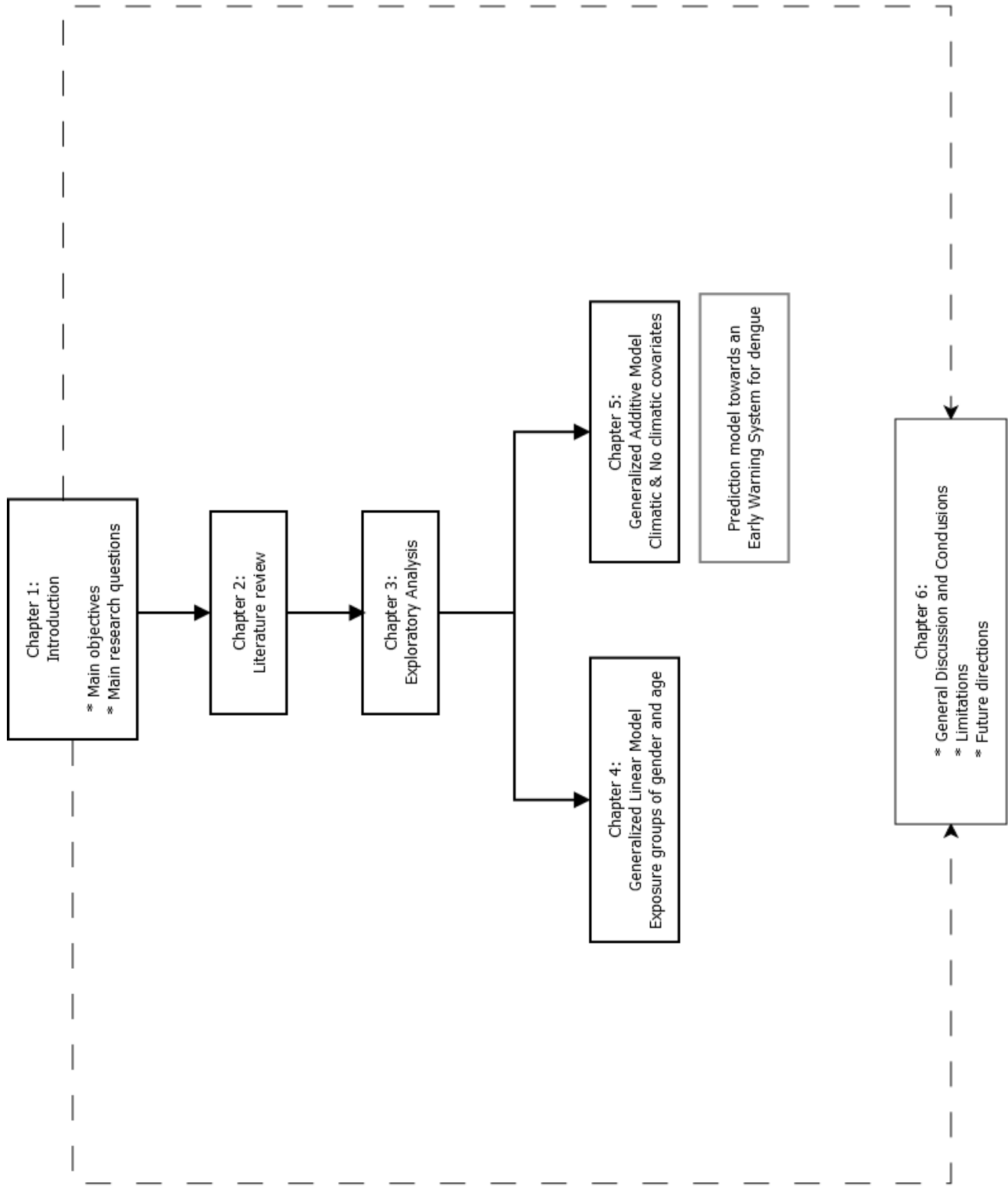


Figure 1.2: Overview of the thesis

Chapter 2

Theoretical background

2.1 Introduction

This chapter provides a general review of the most common statistical methods applied to epidemiological studies in the public health sector. Challenges and limitations which need to be addressed before modelling purposes, have been discussed in detail in this chapter. Special attention was devoted to issues of handling a georeferenced dataset from Zulia state, Venezuela within time and space dimensions. The information implemented in this study was for the first time provided by different sources and at different units.

In addition a framework of explanatory factors associated with dengue transmission has been presented in detail, based on scientific evidence obtained in different parts of the world. Finally, a start point describes the process involved within the literature reviews in conjunction with some key words that merited attention while this project was developed.

2.2 Literature review

The process for seeking information passed through various stages during the development of this study. At an early stage, the review was based on a simple randomized search of references. Titles related to dengue fever comprise the initial findings using the Google Scholar tool. However a more systematic approach was progressively adopted by using more specialized database (Racloz et al., 2012). Academic sources of information were identified in the field of public health and mathematical science with application in statistical methods for epidemiological studies. Because of the nature of this study, the search strategy was conducted following four different stages (from top to bottom), as illustrated in the following diagram (see Figure 2.1).



Figure 2.1: Search strategy on the literature review of this project

It is seen that the top of the pyramid refers to studies of dengue fever worldwide and in a broader fashion. The key words at this stage were: dengue fever, vector borne

disease, *Aedes aegypti*, dengue world, dengue Americas, dengue Venezuela, dengue Zulia amongst others. The second stage involved some clinical and biological implications on dengue transmission. The key words at this stage were: symptoms, vaccine, control, dengue prevention, serotypes, immunity, surveillance and virus evolution amongst others.

These two stages of the pyramid were sought by using the specialized database Pub Med and Web of Knowledge. Based on this process the most common journals read can be arranged according to the impact factor (IF) in the following list: PLoS Medicine, Dengue Bulletin, Environmental Research letters, BMC Public Health, International Journal of Environmental Research and Public Health, Pan American Journal of Public Health.

Subsequent steps entailed a more systematic searching through more technical terminology. Databases such as: Web of Knowledge and MathSciNet were used as main tools for searching information. As a result, the third level of the pyramid in Figure 2.1 named Factors associated with dengue fever, included key words such as: climatic factors and dengue fever, socioeconomic factors (and dengue fever), geographic factors (and dengue fever), demographic factors (and dengue fever) and individual factors (and dengue fever).

Finally the base of the pyramid was reached by including the statistical foundation of the methodologies applied. The most relevant key words at this stage included: spatial regression, Generalized Linear model (GLM), Poisson GLM, Negative Binomial GLM, Generalized Mixed model (GLMM), overdispersion, Spatio-Temporal modelling, Hierarchical Bayesian, disease mapping, prior distribution, WinBUGS, Generalized Additive model (GAM), Generalized Additive Mixed model (GAMM) and Early Warning System amongst others.

As a result a wide range of journals were examined in these latter stages. According

to the IF the journals read at this stage are listed as follows: The Lancet, Environmental Health Perspectives, Emerging Infectious Diseases, PLoS Neglected Tropical Diseases, Tropical Medicine & International Health, American Journal of Tropical Medicine and Hygiene, Occupational and Environmental Medicine, Statistics in Medicine, Biometrics, Computers in Biology and Medicine, Statistics in Medicine, Statistics and Computing and Computers & Geosciences amongst others.

2.3 Surveillance and control of infectious diseases

One of the urgent needs demanded by the public health sector is the endowment of a comprehensive surveillance system with high temporal and spatial resolution, allowing the construction of accurate modeling based on the dynamics of the diseases (Elliott and Wartenberg, 2004). The use of a robust surveillance system is a powerful tool that maximizes the effectiveness of control and prevention programs at local and global levels (Pfeiffer, 1996; Rothman et al., 2008). In particular within unfavoured countries in terms of social welfare (Zacarias and Andersson, 2011; Lowe et al., 2011).

Researchers in the field of spatial epidemiology emphasize that an ideal surveillance system is a network based on Geographical Information System (GIS) in conjunction with remote satellite information for environmental variables (Elliott and Wartenberg, 2004; Rothman et al., 2008). According to the authors the inclusion of advanced statistical techniques and the availability of epidemiological and entomological data (when it is required). Moreover a set of climatic, demographic, socioeconomic and genetic indexes, geographically referenced could be the optimal situation.

However from a practical point of view, some issues handling spatial data have been pointed out:

Lack of appropriate data set.

Diversity in data collection may produce a discrepancy on the quality of records in terms of consistency in physical units at geographical locations over time (Rezaeian et al., 2007; Jarup, 2004). This pattern is mainly observed in routine data provided from health agencies from developing countries which lack geo-referenced information or poor resolution of the data at subregional level (Pfeiffer et al., 2008; Rezaeian et al., 2007; Elliott and Wartenberg, 2004; Chang et al., 2009; Freitas et al., 2003; Rothman et al., 2008).

In this context a case study was found in the northeastern part of Brazil with under-reported cases of dengue fever (Barreto et al., 2008), due to confounding symptoms at the very early stage of the disease. Despite the increased demand for epidemiological data with high resolution. Pursuing statistical studies with advanced techniques becomes a challenge to be overcome in many countries because of the lack of optimal data (Almeida et al., 2009).

As a result, one of the challenges ahead in the health sector must be the consolidation of a robust disease surveillance system in many parts of the world (Elliott and Wartenberg, 2004; Pfeiffer, 1996).

In addition some epidemiological studies have to handle missing climatic data owing to the absence of continuous records provided by meteorological stations. As in the case of an epidemiological study of malaria conducted in Mozambique (Zacarias and Andersson, 2011) or another study performed in Costa Rica (Fuller et al., 2009) implementing an Early Warning System involving dengue transmission. In the case of Mozambique, the missing environmental data was supplied via an interpolation process at point level (Zacarias and Andersson, 2011). Whilst the study conducted in Central America preferred the acquisition of a remote sensing data (Fuller et al., 2009). This last practice has been spread globally, at low-cost to the scientific community (Nakhapakorn and Tripathi, 2005; Lowe et al., 2011).

On the other hand, experts in disease mapping argue about the optimal size of the spatial units. Although some authors claim that disease mapping can be better represented with larger georeferenced units (Rezaeian et al., 2007) due to instability in the response variable, when using lower measures at the geographical level. Others state that the optimal decision needs to consider number of regions large enough to ensure stability in the estimations, but small enough to ensure homogeneity on the risk factors involved within the analysis (Lawson et al., 2004).

Uncertain time delay of the disease. This phenomenon is particularly remarkable in epidemiological studies of infectious vector borne diseases. This approach needs to consider optimal delay time between climatic factors and incidence rates of the disease (Rothman et al., 2008). Scientific evidence reveals a strong impact of climatic factors on the biological cycles of many infectious diseases (Rodier et al., 2004). In particular it has been noticed in dengue fever transmission (Lowe et al., 2011; Chen et al., 2010; Johansson et al., 2009; Hii et al., 2009; Fuller et al., 2009; Hu et al., 2010; Yu et al., 2011).

As a result a comprehensive analysis is required to identify optimal delay time which best captures the dynamics in the transmission of the disease based on different geographical contexts. To this end a theoretical review was debated in section 2.5.2 with a listing of lag times associated with dengue fever in several geographical contexts. Meanwhile an empirical study is conducted in Zulia state in subsequent analyses.

Flexibility of epidemiological information. The fulfillment of this procedure will only be possible once the surveillance process is automated (Rothman et al., 2008). However, there are cases where the process of automating epidemiological data is still in an early stage of development. Such is the case of Zulia state where manual records still exist, despite a great effort to automate that. The creation of a network for surveillance would provide the opportunity to joint efforts amongst health agencies at national and international level to achieve shared goals (Pfeiffer, 1996; Rothman et al., 2008).

2.3.1 Quality and validity of spatial data

Historically the epidemiological data used for public health research has typically been focused on the attributes of the disease. However, the increasing availability of spatially referenced health outcomes based on geographical locations, such as: points, lines or small areas has allowed the implementation of spatial epidemiological analysis which provides insights into causation at geographical level (Pfeiffer, 1996; Pfeiffer et al., 2008; Boscoe et al., 2004).

Generally speaking, geospatial data at point level is defined in terms of latitudinal and longitudinal coordinates but the areal referenced information covers specific geographical entities (Werneck, 2008). In spite of the main interest in the present project being focused on aggregated data at municipality level (small area data available in Zulia state). Some studies of point data modelling dengue risk linked with home address has been illustrated in different parts of the world (Bohra and Andrianasolo, 2001; Cordeiro et al., 2011; Siqueira et al., 2008). But some other studies carried out, have implemented small area as geo-referenced analysis of dengue fever (Freitas et al., 2003; Almeida et al., 2009; Lowe et al., 2011).

On the other hand, although an extensive literature exists concerning epidemiological studies in the field of dengue fever (Lozano et al., 2008; Lowe et al., 2011; Almeida et al., 2009; Hu et al., 2010; Sierra et al., 2007). Some attention has been targeted to entomological spatial analysis using mosquito population sources (Chang et al., 2009; Cordeiro et al., 2011). Both techniques can provide useful findings for the surveillance of dengue fever.

2.3.2 Georeferenced data of dengue fever in Zulia state

The territory of Venezuela is divided into states, which, in turn are divided into municipalities, these being further sub-divided into parishes. A visual representation of the political map of Venezuela with state boundaries was displayed in Figure 1.1 (see

chapter 1). And the map of Zulia state with its municipality boundaries is presented in Figure 2.2.

Zulia state is located on the north west part of Venezuela, geographically located between 8.20° to 11.79° North latitude and between 70.73° to 73.37° West longitude. It is divided into 21 municipalities, over an area of 50,230 square kilometres surrounding the Lake of Maracaibo, the largest lake in Latin America covering 12,870 square kilometres. The 21 municipalities are: Padilla, Paez, Mara, Lossada, Maracaibo, San Francisco, Urdaneta, Santa Rita, Cabimas, Simon Bolivar, Lagunillas, Valmore Rodriguez, Baralt, Miranda, Rosario, Machiques, Semprun, Catatumbo, Colon, Pulgar, and Sucre. According to the National Institute of Statistics, the estimated population in Zulia during 2008 was 3,687,332.



Figure 2.2: Zulia state with its municipalities boundaries: (1) Padilla, (2) Páez, (3) Mara, (4) Lossada, (5) Maracaibo, (6) San Francisco, (7) Urdaneta, (8) Santa Rita, (9) Cabimas, (10) Simón Bolívar, (11) Lagunillas, (12) Valmore Rodríguez, (13) Baralt, (14) Miranda, (15) Rosario, (16) Machiques, (17) Semprun, (18) Catatumbo, (19) Colón, (20) Pulgar, (21) Sucre

The time period chosen for modelling purposes was from January 2002 to December 2008 (364 weeks or 84 months). The geographical entities for the spatial study were selected at municipality level partly owing to the availability of the data, but even more due to the lack of comprehensive consecutive records at smaller scale such as parishes. To this end this research brings together for the first time a large dataset from different sources of data available for Zulia State, Venezuela. These sources of information are detailed as follow:

Epidemiological data:

Hospital admissions of dengue fever and dengue haemorrhagic cases comprise the epidemiological data provided by the Ministry of Health through the local agency in Zulia state. The daily reported cases of dengue fever were obtained at municipality level, from January 2002 to June 2010. However weekly, monthly and annually compiled epidemiological data was aggregated for subsequent studies.

Historical epidemiological records in Zulia state were validated as accurate to pursue modelling studies of dengue fever. According to the local agency of the Ministry of Health, the recorded cases of dengue and dengue haemorrhagic fever followed a unified criterion in all the health centres placed in Zulia, responsible in collecting and recording epidemiological information. This criterion followed the standard clinical symptoms of fever lasting up to seven days, retro orbital pain, in conjunction with headache, abdominal pain and subsequent bleeding symptoms with reduction in platelet counts less than $100,000\text{ }x\text{mm}^3$, following the specifications given by the Ministry of Health in Zulia state (Ministerio, 2007). In addition every declared case belongs to the corresponding municipality avoiding ambiguity or suspected double counting record. However uncertainties regarding type of serotypes circulating during the period of time under analysis were a restriction, which need to be overcome in the future.

More over, the daily reported cases were also stratified by groups of age and gender into the whole Zulia. Unlike the epidemiological data aggregated at municipality level,

had the restriction of age groups and sex in their setup.

Demographic data:

Annual population size aggregated at municipality level was provided by the Ministry of Health through the local agency in Zulia state. This data set covers from 2002 to 2010 and the estimation was based on the first month for each year. The annual estimations, need to be interpolated to fill weekly and/or monthly gaps for the missing information in case it is required in further analyses.

Geographic data:

Average altitude expressed in metres above sea level was obtained from the Ministry of Environment in Zulia state at municipal level. This data, although valuable, has the restriction of using a single value, for the geographical feature throughout the whole municipality.

In addition Zulia state is divided into five main region named: Guajira, East Coast, Capital, Perija and South. Each of these regions comprise an administrative division characterized by having similar socioeconomic development (Alvarez, 2009). This arrangement would be useful for further analysis.

Socioeconomic data:

A set of socioeconomic data geographically geo-referenced at municipality level was provided by the National Institute of Statistics from the Census carried out in Venezuela in 2001. Although the latest national Census was conducted in 2011, the outcomes were not posted when this investigation was developed, so it could not be used.

A wide variety of variables were provided in this category at municipality level. They include: percentage of households with access to the piped water supply network, percentage of households that receive garbage collection service, percentage of unoccu-

ped households, percentage of households with domestic air conditioning equipment, percentage of households including a bathroom without affordable shower, percentage of households that receive electricity service and average of persons living in the same house.

Climatic data:

The historical local weather conditions were initially provided by the Ministry of Environment based in Zulia state. Daily climatic records were obtained via two local stations named: "El Carbon" located on 11.07° latitude and 72.24° longitude and "Machiques" located on 10.05° latitude and 72.56° longitude at 50 metres and 99 metres above sea level respectively. However, an assumption of similarity in meteorological patterns across Zulia state is not reasonable to support the use of these two stations to represent the whole state. As a result the data was rejected owing to not only the small amount of stations but also the missing records. The lack of maintenance on the equipment could be the main cause of failure to continually generate data.

However, recent epidemiological studies of infectious diseases have greatly benefited from the innovative techniques using remotely sensed data (Beck et al., 2000; Lowe et al., 2011). In particular the use of free satellite imagery software such as Google Earth and Geographical Information Systems (GIS) (Eisen and Lozano-Fuentes, 2009; Chang et al., 2009; Lozano et al., 2008) which have proved to be useful tools in developing countries where reliable data and other georeferenced data are not well established. Furthermore international agencies such as the National Oceanic and Atmospheric Administration (NOAA), the French Systeme Pour l'Observation de la Terre (SPOT) and the National Aeronautics and Space Administrations's (NASA) have joined forces to supply the scientific community with very high resolution climatic data.

This project was benefited from the scientific effort provided by the National Aeronautics and Space Administration (NASA) (Acker and Leptoukh, 2007) by implementing a set of remote sensing climatic data over the geographical zone: -74° West to -70°

East and 12° North to 8° South covering the whole of Zulia state, Venezuela. These data were part of the Geospatial Interactive Online Visualization And aNalysis Infrastructure (Giovanni) project based on gridded averaged values within the grids of dimensions 2.0° latitude by 2.0° longitude. In addition the data collected daily were monthly averaged from January 2002 to June 2010. As a result the estimated climatic parameters were computed based on an interactive computational algorithms, which added certain degree of errors and uncertainly. Furthermore parameters of maximum, minimum and average surface temperatures, initially expressed in °K were transformed into in °C. Moreover, maximum, minimum and average rainfall rate expressed in $Kg/m^2/s$ was transformed into the physical unit mm for a better comprehension of the outcomes.

Consequently a new data set with higher resolution pertaining to climatic factors was provided by the National Aeronautics and Space Administrations (NASA) (Kempfer, 2013). The inclusion of the new dataset was owing to the restriction of the initial data provided by meteorological stations in Zulia state due to missed values of continuous records. As a result monthly climatic factors of average, maximum and minimum surface temperature (°C) in conjunction with average, maximum and minimum rainfall rate (mm) aggregated from January 2002 to June 2010 were allocated into four symmetric and non-overlapping grids of dimensions 2.0° latitude by 2.0° longitude covering the whole of Zulia state. The four grids were geographically positioned as follows:

GRID I	−74.0° West to −72.0° East	12.0° North to 10.0° South
GRID II	−72.0° West to −70.0° East	12.0° North to 10.0° South
GRID III	−74.0° West to −72.0° East	10.0° North to 8.0° South
GRID IV	−72.0° West to −70.0° East	10.0° North to 8.0° South

By ensuring the existence and uniqueness of climatic data for each municipality, each capital was placed within a grid according to the latitude and longitude coordinates. Hence the 21 municipalities were all arranged into the four grids as shown in Table 2.1.

Municipality	Capital	Lat-Long	Grid
Rosario	Villa del Rosario	10.32/72.32	I
Machiques	Machiques	10.06/72.54	I
Padilla	Carlos (island)	10.99/71.61	II
Paez	Sinamaica	11.09/71.87	II
Mara	San Rafael	10.97/71.76	II
Urdaneta	Concepcion I	10.42/71.69	II
Maracaibo	Maracaibo	10.63/71.63	II
San Francisco	San Francisco	10.51/71.69	II
J.E.Lossada	Concepcion II	10.62/71.84	II
Simon Bolivar	Tia Juana	10.26/71.36	II
Lagunillas	Ciudad Ojeda	10.20/71.29	II
Cabimas	Cabimas	10.40/71.44	II
Sta. Rita	Sta.Rita	10.52/71.48	II
Miranda	Los Puertos	10.67/71.51	II
J.M.Semprum	Casigua	8.75/72.50	III
Catatumbo	Encontrados	9.07/72.25	III
Colon	San Carlos	9.01/71.91	IV
F.Pulgar	Pueblo Nuevo	8.95/71.63	IV
Sucre	Bobures	9.24/71.16	IV
Baralt	San Timoteo	9.79/71.07	IV
V.Rodriguez	Bachaquero	9.96/71.12	IV

Table 2.1: Geographical location of municipalities by Grid I, II, III and IV

Finally a summary of information provided by the local agencies in Zulia state are detailed above. This data also includes the respective sources (see Table 2.2).

Epidemiological data						
Variable	Description	Scale	Period	Level	Source	
Dengue cases	Hospital admissions	Daily	Jan 2002-Dec 2008	Municipality	Division of Epidemiology, Zulia state. Ministry of Health Idem source (above) Idem source (above)	
Dengue cases	Hospital admissions	Daily	Jan 2009-Jul 2010	Municipality		
Dengue cases	At age and gender	Daily	Jan 2002-Dec 2008	Zulia state		
Climatic data						
Variable	Description	Scale	Period	Level	Source	
Av.temperature	local climatic data	Daily	2002-2008	Stations of Carbon and Machiques	Ministry of Environment based in Zulia	
Min.temperature						
Max.temperature						
Rainfall						
Av.relative humidity						
Min.relative humidity						
Max.relative humidity						
Socioeconomic data						
Variable	Description	Scale	Period	Level	Source	
Proportion of houses with: electricity water supply garbage collection Av. persons in single house	Census data	Annual	2001	Municipality	National Institute of Statistics	
Demographic data						
Variable	Description	Scale	Period	Level	Source	
Population	Population size	Annual	2002-2010	Municipality	Division of Epidemiology, Zulia state. Ministry of Health	
Geographic data						
Variable	Description	Scale	Period	Level	Source	
Altitude	Average altitude			Municipality	Ministry of Environment based on Zulia	

Table 2.2: Initial data set georeferenced in Zulia state provided by local agencies

Missing values and outliers:

Initial data were collected by liaising the Ministry of Health and Environment in Zulia state, Venezuela and spatio-temporal information from a varied of data sources were provided. However as the researcher developed the skills for handling spatial epidemiological studies, some of the data were replaced by more accurate information which supplied missing information. The use of a variety of sources also produced a discrepancy in terms of spatial and temporal units which required to be unified. In addition, some technical issues handling missing data were faced in this study, particularly regarding the epidemiological data. Few gaps spaces were filled with the immediately consecutive daily value when it was needed and consequently the daily records were weekly and monthly aggregated. It was also noticed that manual epidemiological records still exist as main source of information in Zulia state, despite a great effort to automate it. A step forward is needed involving implementing automated records of dengue cases in the public health sector. But more relevant might be the possibility of getting easier access to information for scientific purposes, particularly to accurate information in time and space. Another important issue handling data from Zulia state was the outlier values. At this respect an extensive assessment for exploration was made, aimed at understanding the underlying patterns of the variables (Pfeiffer et al., 2008; Zuur et al., 2010). However although some outliers were observed within the epidemiological dataset over time, these trends were kept because of the importance to analyse in greater detail the evolution of the disease. This allowed the generation of hypotheses which were developed in this present study.

2.4 Statistical methods for epidemiological studies

Spatial epidemiology analyzes the geographic variation in the risk of the disease by finding causes and effects at geographical level (Elliott et al., 2000). This emerging discipline has been enriched by a wide range of statistical techniques aimed to describe, detect, assess and predict the risk of the disease in spatial contexts (Pfeiffer et al.,

2008; Elliott et al., 2000; Elliott and Wartenberg, 2004; Ostfeld et al., 2005; Abellan and Best, 2008). To this end a description of the most common methods applied to epidemiological studies in the public health sector were discussed in this section. However, not all of the methods were applied to the dataset for Zulia state. Nevertheless the theoretical understanding obtained may lead to new ideas and to new topics of interest for further studies.

The epidemiological studies require at first instance an extensive assessment for data exploration, aimed to confirm or provide evidence of patterns that might require a deeper explanation (Pfeiffer et al., 2008; Zuur et al., 2010). To this end a preliminary exploratory analysis needs to be performed to search out patterns on the data before moving on to the construction of more robust techniques in the field of spatial epidemiology.

Towards this goal, experts have promoted the dissemination of a protocol for data exploration to address issues handling future approaches (Zuur et al., 2009, 2010; Zuur, 2012). As a result a summary of techniques for reliable data exploration, is shown in Table 2.3.

Data Exploration	
Protocol	Tools
Outliers (before fitted model)	Boxplots
Collinearity (before fitted model)	Scatterplots and correlation test
Response vs. covariates (before fitted model)	Multi-panel

Table 2.3: Protocol for data exploration including techniques before fitted models

According to the authors the outliers data need to be identified before fitting a regression model. The lack of assessment might lead to overdispersion issues when dealing with more advanced techniques such as Generalized Linear models (GLM) (Cameron and Trivedi, 1999; Zuur et al., 2009; Zeileis et al., 2008). At this respect the box plots are the most common tool to visualize outliers patterns on the dataset.

High collinearity amongst variables, might affect the standard error on the estimations and consequently the p-value which measures the level of significance (Zuur, 2012; Zuur et al., 2009). In this context, a collinearity assessment needs to be performed for selection of covariates, before fitting a regression approach. The most common technique to visualize potential anomalies of multicollinearity is via multipanel scatterplots, tested with the Pearson coefficients (Zuur et al., 2009, 2010; Zuur, 2012).

In addition, the association between response variable and potential explanatory variables might be better represented by multipanel plots which can highlight relevant trends, according to the authors.

2.4.1 Generalized Linear model (GLM) for count data

Aetiological analysis in the epidemiological context is aimed to assess the variation on the health outcome by identifying causation that could explain the distribution of the disease (Elliott and Wartenberg, 2004; Pfeiffer et al., 2008). In this respect, experts on Spatial Epidemiology point out that environmental, social, demographic, genetic and even individual characteristics are the most prominent reasons of causation in public health studies (Wakefield, 2004).

The rapid growth of the use of the methodology Generalized Linear models (GLM) in a variety of fields is due to the flexibility of the approach which allows the inclusion of an extensive set of distributions belonging to the exponential family (Zuur et al., 2009). Experts agree on the use of Poisson GLM when dealing with small and non-negative count data (Pfeiffer et al., 2008; Lovett and Flowerdew, 1989; Schwartz, 1993). Logistic regression is also available when dealing with binary count data in which the response variable might be presence or absence of the disease (Pfeiffer et al., 2008). And in a more general case, the Gaussian linear model when using continuous data (Zuur et al., 2009).

The main purpose in this section is the discussion of the various ways available when handling GLMs with Poisson distributed data. To this end an overview of these techniques (Cameron and Trivedi, 1999; Zeileis et al., 2008) has been summarized in Table 2.4.

Type	Distribution	Method	Features
GLM	Standard Poisson	MLE	variance=mean
	Quasi-Poisson	Quasi MLE	$var > mean$ or $var < mean$
	Negative Binomial	Quasi MLE	$var > mean$ or $var < mean$
	Zero Inflated	MLE	More zeros than expected on the response
	Zero Truncated	MLE	Zero value is not allowed
	Mixed Poisson	MLE	Extension of Poisson-GLM using random effects

Table 2.4: Overview of parametric Poisson GLM for count data

The simplest of these techniques is standard Poisson distribution with an estimation process based on maximum likelihood function (MLE). As can be seen in Table 2.4, the mean and the variance using the standard GLM Poisson approach are the same values. However this approach has been avoided by most of the experts because it results in an unrealistic pattern in ecological and epidemiological studies (Lowe et al., 2011; Cameron and Trivedi, 1999; Zeileis et al., 2008), which can lead to false estimations.

However in a more realistic situation the variance might exceed the mean, leading to overdispersion or might be less than the mean, leading to underdispersion (Zuur et al., 2009). Both abnormalities are attributed to: unknown or unobserved covariates, confounding variables, model misspecification when important interaction terms are missed or when the correct form of the relation between dependent and independent factors is omitted (Zuur et al., 2009; Waller et al., 1997).

Owing to these phenomenon, tests to measure the dispersed parameter are recommended to justify the use of more complex techniques, which enables more realistic estimations (Cameron and Trivedi, 1999; Zuur et al., 2009). In such circumstances, the alternative methods of Quasi-Poisson and Negative Binomial accounting for overdis-

persion are available (Zuur et al., 2009).

Experts argue that both approaches, although with the same aim, might produce a massive difference in the estimated parameters (Hoeff and Boveng, 2007). So, the selection of the best approach is essential to ensure reliable findings. In the Quasi-Poisson approach for instance, the variance is defined as a product between the overdispersion parameter and the mean (Hoeff and Boveng, 2007; Zuur et al., 2009; Cameron and Trivedi, 1999). Another approach that takes into account the heterogeneity of the variance is the Negative Binomial (NB). This in turn, defines the overdispersion parameter as a multiplicative factor which depends on the mean (Zuur et al., 2009).

From a pragmatic point of view, some authors suggest the estimation of dispersion parameter based on the ratio between the Pearson's Chi squared and the degrees of freedom (Zuur et al., 2009). Hence, the dispersed parameter ϕ can takes the following options:

- If $\phi = 1$** variance equal to mean leading to standand Poisson
- If $\phi > 1$** over-dispersion
- If $\phi < 1$** under-dispersion

Although some studies seem to prefer the use of Quasi-Poisson approach as the best fitted model(Hoeff and Boveng, 2007) when dealing with ecological applications, others prefer the Negative Binomial approach in clinical investigations (Schukken et al., 1991). However experts in applied statistics suggest as the best option, the use of Negative Binomial when the dispersion parameter is greater than 15 (Zuur et al., 2009).

Using the Quasi-Poisson or Negative Binomial GLM approaches in epidemiological context, some authors have implemented the use of an offset term (Lowe et al., 2011). This term has been based on a log of the population within a particular region, in order to adjust for differences in the population size across the underlying regions when dealing with rates (Terzi and Cengiz, 2009).

According to the summary presented in Table 2.4, some other options in modelling count data with a parametric framework are the Zero Inflated Poisson model (ZIP) based on more zeros than expected on the response variable (Zuur et al., 2009; Cheung, 2002; Cameron and Trivedi, 1999). Or the Zero Inflated Negative Binomial (ZINB) which deal with extra-variation on the Poisson structure (Cheung, 2002; Zuur et al., 2009). These options though lacking of interest in the current study because of time, can be found in detail, in the cited references.

Nevertheless, the issues might arise when more complex structures are required for Spatio-Temporal modelling. In the case of Generalized Linear model (GLM) the overdispersion occurs at the individual level of the observations (Hara, 2009), and the main purpose being the comparison amongst one group against the other. However when new extravariation occurs at different levels of the analysis, an extension of the Generalized Linear model (GLM) into a Generalized Linear Mixed model (GLMM) (Gelman and Hill, 2007) is suggested.

As a result the new structure of GLMM comprises fixed term and an extra random effect to account for the heterogeneity of the variance (Zuur et al., 2009). On the one hand the fixed term includes all the covariates within the model whose impact on the response outcome is fixed (Mullen and Birkeland, 2008; Zuur et al., 2009). On the other hand the random term allows for the real random part accounting for the overdispersion of the data, and the term of random effects in the form of temporal and/or spatial dependencies (Zuur et al., 2009).

2.4.2 Generalized Linear Mixed models (GLMM) for count data

The Generalized Linear Mixed models (GLMM) have grown as an extension of GLMs in a variety of applications (Zuur et al., 2009; Mullen and Birkeland, 2008; Gschlobl and Czado, 2008). Although several authors use different names to denote it, such as: Mixed models, Hierarchical linear models or Multilevel analysis, the methodology re-

mains the same (Zuur et al., 2009). Epidemiological and ecological studies stand out a great benefit of using GLMM, since the inclusion of heterogeneity within the variance, through the use random effects accounting for spatial and/or temporal dependencies (Zuur et al., 2009; Lowe et al., 2011; Das et al., 2002).

Often the data is correlated at spatial and temporal scale, in which case a Generalized Linear Mixed approach has been pointed out to be the best technique (Almeida et al., 2009; Lowe et al., 2011). A study conducted in Brazil (Lowe et al., 2011) promotes the benefits of this approach via a Spatio-Temporal modelling of dengue fever using a Negative Binomial GLMM in a Bayesian fashion.

Although some methods are available to estimate the parameters of a GLMM structure, the implementation of a Hierarchical Bayesian framework, has been one of the most common approaches used in epidemiological studies (Das et al., 2002; Lowe et al., 2011). From a practical point of view the fixed and random terms via a Hierarchical Bayesian perspective are treated as random with appropriate prior distributions (Das et al., 2002; Lowe et al., 2011). And the computational implementation is most easily achieved by using the free software WinBUGS (Lowe et al., 2011; Zuur et al., 2009).

Spatial autocorrelation

Spatial and temporal autocorrelations need to be assessed to justify the application of a more complex structure of GLMMs. To this end the spatial autocorrelation is based on Tobler’s first law of geography, which states that measured values located in nearby regions tend to exhibit more similar values compared with values far apart. The Moran’s I is the most common method of measuring such spatial association (Stevenson, 2011). This measure has been defined as a linear association between the observed data and the weighted average, by taking into account the surrounding areas (Stevenson, 2011). Hence the Moran’s I formulation is as follows:

$$I = \frac{n \sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2 \left(\sum_{i \neq j} \sum w_{ij} \right)}$$

where n is the number of polygons, w_{ij} the value of the defined matrix of proximity between observation i and j . The y_i the health outcome and \bar{y} the sample means of the health outcome.

The criterion of independence of data is commonly missed in spatial analysis due to the effect of unknown or unobserved factors involved in the analysis (Lawson, 2009; Banerjee et al., 2004; Lowe et al., 2011). These factors can be included within the hierarchical Bayesian structure with an appropriate prior distribution, which in conjunction with the available dataset or likelihood function produces an entire posterior distribution (Lunn et al., 2000; Gimenez et al., 2009; Osei, 2010; Lawson, 2009). As a result hierarchical levels based on the Bayesian framework are obtained. In some of the cases the selection of the prior distribution of the parameters is vague, flat or non-informative, resulting in a posterior distribution that strongly depends on the likelihood of the data (Lawson, 2009).

Another concern dealing with spatial data using spatio-temporal modelling is how to define the adjacency matrix. This matrix needs to identify the associations between the areas involved on the geographical representation. Although various options are available, the method selected depends on the nature of the dataset (Stevenson, 2011). According to the author the criterion in defining such matrix can be based on:

- Areas having borders in common
- Areas having corners in common
- Areas having centroids at specific distance
- Areas having centroids at inverse or squared inverse specific distances

The recent implementation of GeoBUGS and WinBUGS packages in the context of Hierarchical Bayesian analysis have contributed in defining adjacency matrix (Lawson et al., 2003; Solymosi et al., 2010; Thomas et al., 2004). As a result, the adjacency matrix can be automatically defined via WinBUGS, after importing the maps using the GeoBUGS tool.

Temporal autocorrelation

The independence of the residuals over time is another statistical assumption that needs to be assessed (Zuur et al., 2009; Zuur, 2012). The most common plots being the autocorrelation function (ACF) and the partial autocorrelation function (PACF) (Zuur et al., 2009; Zuur, 2012) in which the temporal dependency can be observed. The violation of such assumption leads to the inclusion of a temporal autocorrelation structure within the model (Zuur et al., 2009; Zuur, 2012).

In the case that an exponential decay is observed along the ACFs and PACFs plots, an autoregressive (AR) structure of order 1 seems to be the best temporal option to be included (Root, 2011). However some other options such as: compound symmetry structure (Zuur et al., 2009), acknowledged to be the most general version. And the autoregressive moving average (ARMA) which is an extension of the autoregressive of order 1 (Zuur et al., 2009).

2.4.3 Disease mapping

One of the tools in many epidemiological studies is the construction of disease mapping to represent the risk rates of health outcome via a Bayesian perspective (Spiegelhalter, 1998; Wakefield, 2007; Waller et al., 1997; Lawson and Fiona, 2001; Lawson, 2009).

Estimating disease risks using a Bayesian hierarchical framework is usually the most powerful tool which enables understanding of complex structures when the risk varies

at a geographical level, allowing the inclusion of unknown or unobserved quantities and the presence of potential dependencies over time or space (Lawson, 2009; Banerjee et al., 2004). To this end the epidemiological data of observed cases varying over time and space can be denoted by:

$$Y_{it} = (y_{1t}, y_{2t}, \dots, y_{nt})$$

and the vector of expected cases :

$$E_{it} = (e_{1t}, e_{2t}, \dots, e_{nt})$$

The e_{it} is the expected number of cases at geographical unit i and temporal period t . It is proportional to the population at risk, multiplied by the global incidence rate over the whole region at a particular period of time (Martinez et al., 2008). This formulation is described below:

$$e_{it} = p_{it} \times \frac{\sum y_{it}}{\sum p_{it}}$$

As a result, the standardized incidence ratio or standardized mortality/morbidity ratio (SMR) widely used by epidemiologists is defined as the ratio of observed counts and expected counts at geographical level (Lee and Mitchell, 2012; Lawson, 2009), in the following form:

$$(\theta_{it}) = \frac{y_{it}}{e_{it}}$$

So the general specification for modelling relative risk at a hierarchical level, accounting for time and space is:

$$Y_{it} \sim \text{Poisson}(\theta_{it}e_{it})$$

where the \log of relative risk in area $i = 1, \dots, N$ and time $t = 1, \dots, T$ defined in the following additive form:

$$\log(\theta_{it}) = \alpha + \lambda_i + \epsilon_t + v_{it}$$

where α is the overall value representing the intercept term of the model. The main spatial effect λ_i can be broken down into structured and unstructured random effects. The term ϵ_t represents the main time effect capturing the spatial dependence over time, which can also be split into structured and unstructured random effects and v_{it} the term which captures the space-time interactions (Zacarias and Andersson, 2011; Richardson et al., 2004).

In this context, the unstructured random effect of the spatial main effect λ_i allows for extra poisson variation called overdispersion term in most of the epidemiological studies (Lawson, 2009; Banerjee et al., 2004). And the spatial structured random effect accounts for spatial dependence amongst geographical units (Lawson, 2009; Banerjee et al., 2004). Likewise the inclusion of structured random effects in the temporal term allows for temporal dependence and the unstructured accounts for temporal heterogeneity (Richardson et al., 2004).

Although different alternatives to account for spatio-temporal associations are available using the Bayesian perspective (Guha and Ryan, 2006; Martinez et al., 2008; Kim

and Lim, 2010). The CAR model offers the advantage of being computationally more flexible when dealing with spatial data (Guha and Ryan, 2006). According to these authors, it is seen that given the spatial adjacency matrix, the CAR Normal represents the prior information of the spatial effect, and simultaneously, given the temporal adjacency matrix, the CAR Normal represents the prior information of the temporal effect.

2.4.4 Generalized Additive models (GAM) for count data

The parametric Generalized linear model (GLM) and Generalized Linear Mixed model (GLMM) are limiting because of the linear assumption. However a more flexible extension is offered by a Generalized Additive model, first implemented by Hastie and Tibshirani (Hastie and Tibshirani, 1984, 1986). The flexibility of this approach allows the replacement of the linear structure with a set of unknown smooth functions to explain more complex relationships (Terzi and Cengiz, 2009).

In recent times, many followers have acknowledged the advantages of this technique (Lado et al., 2008; Zuur, 2012). In particular it has greatly benefited the applications in Ecology (Murase et al., 2009; Zuur et al., 2009; Zuur, 2012) and Public Health sector (Luo, 2012).

According to experts the benefits of the GAMs (Berg, 2005; Terzi and Cengiz, 2009; Murase et al., 2009; Zuur et al., 2009; Zuur, 2012) can be summarized as:

1. GAM is an extension of GLM via smoother functions that can better capture more complex structures.
2. The parametric assumptions are avoided, which makes this approach a more flexible tool for modelling.
3. The outcomes can be visualized via the smooth curves posed by each explanatory variable.

4. The inclusion of interaction terms amongst covariates can be handled with this approach.

The most general structure(Wood, 2006; Clifford et al., 2011; Terzi and Cengiz, 2009) can be formalized below:

$$g(\mu_i) = X_i^* \theta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}x_{4i}) + \dots$$

where:

X_i^* is the parametric structure

x_{ji} the explanatory covariates

$f_1, f_2 \dots$ the smooth functions

f_3, \dots the interaction term amongst covariates

Y_i , the observed variable

If Y_i are referred to count data, a Poisson distribution is the best option (Zuur, 2012). Several packages are nowadays available to conduct smoothing approaches (Zuur, 2012). The most common being: *gam*, designed by Hastie; *gamlss*, developed by Rigby and Stasinopoulos. But the most widespread tool is the *mgcv*, whose authorship belongs to Wood (Wood, 2006). This package has been pointed out as one of the best options based on its versatility (Zuur, 2012) in terms of: The easy inclusion of spatial and temporal autocorrelation, the identification of the smooth functions based on an automatic process, the random effects can be also included. Finally this package provides a variety of options regarding the smoothing techniques (Wood, 2006) which can be summarized below:

- thin plate (bs= tp), which estimate the smooth functions measuring the wiggleness

- thin plate with shrinkage ($bs = ts$), in which the smooth function cannot be completely zero
- cubic smoothing ($bs = cr$), having integral of the second order derivatives as a penalty
- smoothing with shrinkage ($bs = cs$), the smooth functions can be completely zero
- cubic smoothing for cyclic term ($bs = cc$), having cubic smoothing for cyclic term

Random effects statistical models

A contribution to knowledge made by this dissertation has been the application of a generalized additive mixed model using epidemiological data of dengue fever in Zulia state, Venezuela with municipalities as a random effect (see chapter 5 for more detail). In this respect a random effect structure assumed in this study confirmed that the health outcome of incidence rate across the geographical units vary in a random fashion. Experts have attributed this variability to real differences amongst the geographical units (Riley et al., 2011) which may be partly attributable to unmeasured intrinsic population differences. However some other factors such as differences in percentage between of urban and rural population within different municipalities might be highlighted as another potential cause. Riley suggested that there might be interventions programmes during the period of analysis which have had varied effects between different municipalities (Riley et al., 2011). In addition, there are uncertainties with respect to epidemiological serotypes circulation across the state which might have different effects per municipalities (Bennett et al., 2009). And finally, cases of dengue stratified differently by groups of age and gender within municipality levels might perhaps have some effect in the underlying heterogeneity. Consequently, the followers of random effect techniques argue that more studies are needed to elucidate the factors which cause heterogeneity (Riley et al., 2011). However, according to these authors the difficulty in achieving this goal is due to the enormous complexity in understanding the spread of the disease at geographical level with a lack of reliable information in most of the cases. Consequently random effects statistical models will dominate the field in

applied studies in the public health sector for a while (Riley et al., 2011).

GAM methodology applied to dengue transmission

Although few scientific reports were found using the GAMs technique in modelling dengue fever. Some of them have been based on entomological collected dataset of dengue fever (Cordeiro et al., 2011; Siqueira et al., 2008; Honorio et al., 2009). However its goodness needs to be fully leveraged in terms of the smooth functions and perhaps the implementation needs to be justified.

In the study conducted in Central Brazil (Siqueira et al., 2008) the collected data was via questionnaire and blood test of the participants infected by dengue fever. As a result limitations concerning the reliability information was observed. In addition, the model specification might lead to conflict of multicollinearity amongst covariates because no tests were conducted to assess this issue. Finally only one smooth term was considered in the model, so perhaps more covariates need to be explored in deeper detail.

The case study performed in Rio de Janeiro (Honorio et al., 2009) southwest of Brazil (Cordeiro et al., 2011) in turn followed similar specifications compared with the previous one. Likewise only one smooth function was implemented, thereby missing the richness of the technique by exploring more covariates.

To conclude, a wide spectrum of opportunities is provided for modelling spatial and temporal distribution of dengue fever, by using Generalized Additive models in a non-linear fashion. As a result the goal of this research is to develop a GAMM structure using epidemiological dataset from Zulia state, Venezuela using random effect at municipality level. Appropriate spatial and temporal autocorrelations need to be assessed to justify whether an extended model is required. Finally the GAMM structure will be used for predictive purposes.

The use of GAM for forecasting

The global challenge commissioned by the WHO is the prevention and control of infectious diseases based on information provided by the Early Warning Systems validated with highest standards of efficiency (Chan, 2012). These authors argue that a wide understanding of historical epidemics in conjunction with the underlying factors associated with the diseases is crucial for developing an efficient tool to detect epidemics in advance. Although several studies have contributed to a prompt detection of malaria in different parts of the world (Bouma and Van der Kaay, 1996; Bouma et al., 1997; Thomson et al., 2004; Abeku et al., 2004). Only few studies have turned to the development of dengue forecasting (Lowe et al., 2013; Descoux et al., 2012; Hii et al., 2012; Chi et al., 2012) applying different statistical techniques (see Section 2.6.2).

This section aims to explore the real challenges in the construction of forecasting models based on generalized additive models. However the paucity of studies using this technique became the main obstacle to be overcome. No case study was found in the public health sector using generalized additive forecasting models with epidemiological data of dengue fever. However some useful evidence revealed technical limitations using this methodology for forecasting purposes, particularly to forecast salmon abundance (Wang and Morishima, 2009) in ecological studies, and weather forecasting (Chi et al., 2012; Vislocky and Fritsch, 1995) in the environmental field. From a practical point of view, the generalized additive models are run more slowly when compared to linear models. This leads to disadvantages with large datasets (Vislocky and Fritsch, 1995). According to these authors another computational concern of the GAM structures is the amount of storing space required for the predictive information. Furthermore, a more detailed analysis established the issue in interpreting and applying the predictive outcomes from a generalized additive approach (Wang and Morishima, 2009). These authors suggest the inclusion of parametric models once the shapes of the associations have been identified by the non-parametric models. Thus the combination of both

non parametric and parametric methodologies seem to be the ideal way to construct forecasting models of dengue fever in Zulia state, Venezuela.

2.5 Explanatory factors associated with dengue fever

2.5.1 Individual factors

Epidemiological studies of dengue fever in different parts of the world exploring vulnerability on population groups have provided evidence that age, gender and race affect the severity of the disease (Raja and Devi, 2006; Sierra et al., 2007; Kongsomboon et al., 2004; Morrison et al., 1998). However more studies are needed to properly identify groups at risk for the purpose of directing effective control measures to benefit the most vulnerable population.

According to experts, a study conducted in Cuba about genetic susceptibility to the disease amongst racial groups, revealed that "caucasians" are more likely to get dengue fever, whilst those of black African ancestry are highly resistant (Sierra et al., 2007). These findings come from three severe epidemics which took place on the island during 1981, 1997 and 2001. Although this thesis does not explore this particular risk factor, since the epidemiological data in Zulia state does not record racial group in reported cases, observations in this direction could be very valuable.

Age as a risk factor is more closely examined in epidemiological studies of dengue fever. Much evidence confirms that children are the most vulnerable population for both mortality and morbidity (Kongsomboon et al., 2004; Morrison et al., 1998; Raja and Devi, 2006). In this respect, efforts in reducing mortality of children is the most important priority identified by the United Nations in the Millennium Development Goals report (Ban, 2007).

Gender is also an important factor in epidemiological studies of dengue fever, al-

though some studies undertaken in Latin America revealed no evidence of variation between gender (Guha-Sapir and Schimmer, 2005). As a result a study needs to be conducted in Zulia state to explain the trend in risk amongst gender and group of ages.

2.5.2 Specific serotype circulating

There are four kinds of virus which have been clearly identified as responsible for dengue transmission, called DENV1, DENV2, DENV3 and DENV4. However, the recognition of multiple serotype activity in various parts of the world, has complicated the situation in the public health sector (Halsey et al., 2012). These authors argue that the severity of the disease is strongly associated with infection regarding the serotypes circulating based on a comparative study performed in some Latin American countries (Peru, Bolivia, Ecuador and Paraguay). However they also emphasize the need for more studies to elucidate the severity of the disease amongst different age groups, including children and adults, using the DENV serotype categories.

2.5.3 Climatic factors

Several scientific investigations have established the relation between outbreaks of dengue fever and climatic variations in temperature (Lowe et al., 2011; Yu et al., 2011; Chen et al., 2010), rainfall (Lowe et al., 2011; Chen et al., 2010; Hii et al., 2009) and humidity (Chen et al., 2010) as well as the global phenomenon of El Nino Southern Oscillation (ENSO), which has emerged to aggravate the risk of dengue transmission in some parts of the world (Lowe et al., 2011; Yu et al., 2011; Fuller et al., 2009; Hu et al., 2010; Kovats, 2000; Githeko et al., 2000; Epstein et al., 1998). Thus, these studies have provided strong evidence of the biological susceptibility of *Aedes aegypti* species to different weather scenarios (Kovats, 2000; Hopp and Foley, 2001).

Temperature associated with dengue fever:

Some studies have claimed that warmer temperatures affect the spread of the disease

in various ways. First of all, a higher temperature reduces the larvae size of mosquitoes which turn into smaller adults, that tend to digest blood faster and consequently more blood is required to develop their eggs (Brunkard et al., 2008; Thomas et al., 2006).

Secondly, areas considered inhospitable for the development of the mosquito population may become ideal locations for the flourishing of the species (Kovats et al., 2003; Brunkard et al., 2008). Cases of dengue for instance have been reported in mountain areas located in Asia, Africa and Latin American countries, considered previously to be too cold for mosquito offspring (Epstein et al., 1998; Hopp and Foley, 2001). Furthermore, some researchers have found at certain temperatures, there is a reduction in the time it takes for the female mosquito to become infected once blood is taken from a human being (extrinsic incubation) (Githeko et al., 2000; Brunkard et al., 2008). Consequently, there is more rapid dengue transmission according to the authors.

For example, a study conducted in Mexico (Brunkard et al., 2008) reported that the incubation period drops to seven days when the temperature ranges between 32°C to 35°C as compared to 12 days when it is 30°C.

Rainfall associated with dengue fever:

With regards to rainfall intensity, some scholars have argued that heavy rainfall may destroy the larvae of *Aedes aegypti* mosquitoes (Hii et al., 2009). But although heavy rainfall reduces the numbers of the species in the short term, the abundance of breeding sites may increase the spread of the disease in the long run (Hii et al., 2009).

It has also been suggested that low rainfall levels can lead to a rise in temperature with the result that people increase their usage of water storage containers, which is a practice that has been associated with the spread of dengue fever (Bohra and Andrianasolo, 2001). Other investigators have emphasized the increase in the amount of dense forest cover, as a long term result of intense rainfall (Githeko et al., 2000). Being a suitable habitat for the mosquito population. On the other hand, some other studies

have reported a strong association between high humidity and dengue transmission (Nakhapakorn and Tripathi, 2005; Chen et al., 2010) owing to the longevity of the species.

The El Nino Southern Oscillation (ENSO) associated with dengue fever:

Inter-annual fluctuations in dengue fever outbreaks have been associated with the global climatic phenomenon of El Nino Southern Oscillation (ENSO) (Gagnon et al., 2001; Hales et al., 1999; Kovats, 2000). This phenomenon originates in the Pacific Ocean, starting with fluctuations in the sea temperature, moving from the east coast of Australia to the west coast of South America, causing global alterations in the atmospheric conditions every 2 to 7 years (Kovats et al., 2003; Hales et al., 1999). Moreover, temperature and rainfall anomalies have been associated with the El Nino phenomenon in some Latin American countries (Fuller et al., 2009; Hales et al., 1999). The presence of the El Nino phenomenon has led to severe dengue outbreaks owing to biological evolution in the transmission of the virus (Kovats et al., 2003).

The two extremes of the ENSO are, on the one hand El Nino (the warm event), which consists of a warming of the sea surface temperature and on the other, La Nina (the cold event), where there is cooling.

Researchers who have assessed the impact of the ENSO on dengue transmission have used a variety of indexes, such as: Sea Surface Temperature (SST) (Brunkard et al., 2008; Fuller et al., 2009; Lowe et al., 2011) in which the Pacific Ocean is split into five geographical zones capturing the mean sea surface for each one, named: Nino1, Nino2, Nino3, Nino4 and Nino3.4 (Fuller et al., 2009). Nino3.4 is an overlapping zone on Nino3 and Nino4 (see Figure 2.3).

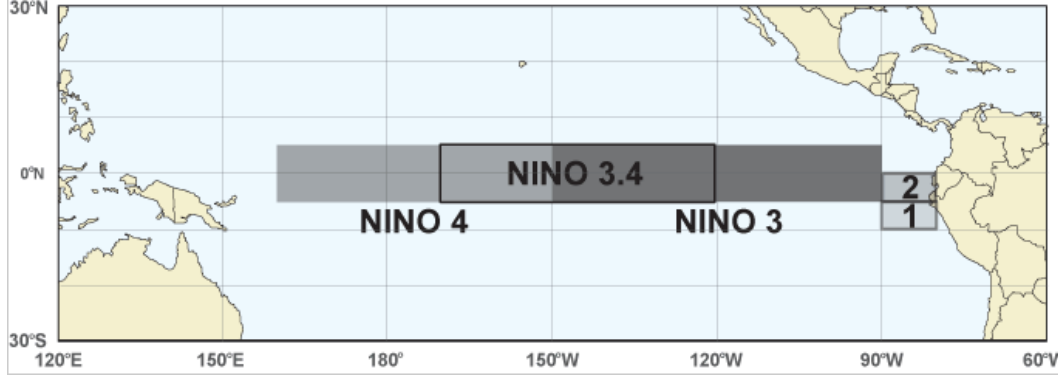


Figure 2.3: Geographical regions of the El Niño Southern Oscillation (ENSO) on the Pacific Ocean

The five zones described in Figure 2.3 are geographically located as follow:

Nino1	5°N to 10°S and 80°E to 90°W
Nino2	0°N to 5°S and 80°E to 90°W
Nino3	5°N to 5°S and 150°E to 90°W
Nino4	5°N to 5°S and 160°E to 150°W
Nino3.4	5°N to 5°S and 120°E to 170°W

A climatic analysis performed in Mexico (Brunkard et al., 2008) using the Sea Surface Temperature (SST) in zone Nino3.4, revealed a significant association with dengue incidence. And another study undertaken in Costa Rica used the SST for predicting dengue epidemics (Fuller et al., 2009). Those findings suggest that El Niño in zone 3 and 4 are more likely to affect countries located in South America.

The findings in both proposals are consistent with the fact that the association among dengue cases and the ENSO- SST index was found negative, meaning that La Nina (cool event) has a stronger effect on the dynamic of dengue transmission in those countries. These results have been confirmed in a recent study performed in Brazil (Lowe et al., 2011) in which the SST in zone El Niño3.4 was found negatively associated with incidence of dengue fever. In this respect the El Niño3.4 better captures the association between the two extremes of the climate, i.e. the warm extreme (El Niño) and the colder (La Nina) (Lowe et al., 2011; Fuller et al., 2009).

In this regard the SST index has been suggested to pursue climatic investigations within the Caribbean area (Fuller et al., 2009). However, the Southern Oscillation Index (SOI) is seen as another parameter for measuring the anomalies of the ENSO, being regarded as a good predictor in epidemiological studies of dengue, undertaken in Australia (Hu et al., 2010) and Taiwan (Yu et al., 2011).

Bearing in mind the above findings, the Nino 3.4 seems to be the best predictor of dengue incidence in the north western part of Venezuela during the period from 2002 to 2008. Endorsed by the fact that the El Nino 3.4 better represents the climatic pattern during the years 2006 and 2007 (McPhaden, 2008) in which the highest epidemic occurred in Zulia state, as referred in this project (see Chapter 3).

In addition, a good understanding of the biological dynamic in infectious diseases regarding climatic variables is a crucial step for epidemiological studies due to uncertainties regarding temporal associations (Rothman et al., 2008; Fuller et al., 2009; Sattenspiel, 2009). Some experts have emphasized the need to consider incubation periods in both mosquitoes and human population, as well as biological cycles of reproduction and the maturation and lifespan of the organism (Hu et al., 2010; Hii et al., 2009). To carry out accurate epidemiological studies regarding lagged effect of dengue transmission, a study performed in Thailand, found that the complete mosquito life cycle, including: egg, pupa and adult stages, ranges from 7 to 12 days (Nakhapakorn and Tripathi, 2005). Other studies have pointed out that two processes are involved with dengue transmission (Henchal and Putnak, 1990), summarized in Table 2.5.

On the one hand, as depicted in the top part of the table, the virus circulating from an infected mosquito to a non-infected human host, which involves an intrinsic incubation, can range from 3 to 14 days (Hii et al., 2009). This is followed by a viraemic period in which the patient manifests the first symptoms of the disease and an onset of fever can be extended for about 6 days, developing later into the worst stage

Infected mosquito to non infected human	
Description	Period (days)
Intrinsic incubation	3 to 14
Viraemic period	2 to 10
Febril period	4 to 6
Haemorrhagic symptoms	1 to 5
Non- Infected mosquito to non infected human	
Description	Period (days)
Extrinsic incubation	8 to 12

Table 2.5: Temporal evolution of dengue transmission in human population

of the disease with haemorrhagic complications (Henchal and Putnak, 1990). Taking the maximum number for each of these stages results in a total figure of 35 days, when those who go on to develop the most serious phase of the disease will have exhibited the symptoms associated with it.

In addition, the virus can pass from an infected human to a non-infected mosquito through an extrinsic incubation period of between 8 to 12 days (Hii et al., 2009), when it becomes infected and when the higher figure is added to the maximum 12 days from birth to adulthood (Nakhapakorn and Tripathi, 2005), this gives a total of 24 days. Consequently, the complete evolution period from hatching to the final disease stage is 59 days or approximately two months. Those theoretical findings are consistent with empirical outcomes seen in Brazil (Lowe et al., 2011), Puerto Rico (Johansson et al., 2009) and Taiwan (Chen et al., 2010) regarding the appropriate time lag for temperature and rainfall associated with dengue transmission. See Table 2.6.

However, as can be seen in Table 2.6, those times vary quite substantially across empirical studies regarding significant time lag for climatic factors, undertaken in different parts of the world. Those variations have been attributed to local weather conditions providing a non-homogeneous pattern regarding the spread of the disease (Yu et al., 2011; Johansson et al., 2009).

In this regard an empirical estimation of relevant delay time associated with cli-

Index	Time lag	Location	Reference
Rainfall (mm)			
Av. rainf (month)	-1 to -2	Brazil	(Lowe et al., 2011)
Rainf (month)	-2	Taipei (Taiwan)	(Chen et al., 2010)
Rainf (month)	-3	Kaohsiuny (Taiwan)	(Chen et al., 2010)
Cum rainf (month)	-1 to -2	Puerto Rico	(Johansson et al., 2009)
Cum rainf(week)	-1 to -20	Singapore	(Hii et al., 2009)
Temperature (°C)			
Av. surf air temp (month)	-1 to -2	Brazil	(Lowe et al., 2011)
Av.temp (month)	-1	Taipei (Taiwan)	(Chen et al., 2010)
Av.temp (month)	-3	Kaohsiuny (Taiwan)	(Chen et al., 2010)
Max.temp (month)	-2	Taipei (Taiwan)	(Chen et al., 2010)
Max.temp (month)	-3	Kaohsiuny (Taiwan)	(Chen et al., 2010)
Max.temp-min. temp (month)	0 to -2	Puerto Rico	(Johansson et al., 2009)
Av.temp (week)	-5 to -20	Singapore	(Hii et al., 2009)
Relative Humidity (%)			
R.H. (month)	-4	Taipei-Kaohsiuny (Taiwan)	(Chen et al., 2010)
ENSO			
Nino3.4 (month)	-6	Brazil	(Lowe et al., 2011)
Nino3.4 (week)	-11	Costa Rica	(Fuller et al., 2009)
SOI (month)	-3 to -12	Australia	(Hu et al., 2010)
SOI(week)	-11	Taiwan	(Yu et al., 2011)

Table 2.6: Climatic predictors with appropriate time lag associated with dengue fever

matic dataset from Zulia need to be clarified in subsequent studies.

By considering each of these studies listed in Table 2.6 in turn, a research carried out in Brazil (Lowe et al., 2011) identified socioeconomic and climatic risk factors based on a Generalized Linear Mixed model (GLMM) of count data. These findings revealed a positive association with temperature and rainfall regarding the transmission of dengue fever. The most significant time lag in this study was identified at 1 and 2 months after rainfall events and temperature fluctuations respectively. Furthermore, the El Nino 3.4 index was found to be statistically significant at negative level after 6 months of occurrence of the event. This study implemented a spatio-temporal Generalized linear mixed model via Bayesian framework to capture the spatial and temporal variation of unknown factors associated with the risk of the disease using structured and unstructured random effects.

According to the investigations listed in Table 2.6, another study in detecting

dengue delay of climatic variables affecting dengue incidence was conducted in Taiwan during 2001 to 2008 (Chen et al., 2010). This approach assumed a linear association between explanatory variables and health outcome. The study was performed with data from the north and the south of the country owing to the severity of the disease. The dengue incidence was modelled using a Generalized Poisson Estimating Equation (GEE) yielding the following results: the delay time for climatic variables involved within the study showed a different values for each region. The most significant at 0 and 1 month for mean temperature before the effect of the disease in Taipei, located on the north part of Taiwan, and significant at 3 months in Kaohsiung, in the south part, linked to mean temperature. On the other hand, the maximum temperature was also assessed in Taiwan revealing the most significant delay time at 2 months in the north, compared to 3 months in the south of the country. Moreover rainfall levels also showed a dissimilar patterns amongst north and south regions. On one hand a delay time of 2 months in the north and 3 months before dengue events on the south of Taiwan.

These findings confirm the statement claiming that delay time estimation for climatic factors regarding dengue transmission might vary from place to place (Johansson et al., 2009) even in regions located in the same country as it was showed above. Hence the importance of discerning an empirical association of dengue incidence and weather conditions in Venezuela becomes a priority.

From references in Table 2.6, some studies assumed a non-linear association between climatic factors and dengue transmission. These approaches have been supported by studies undertaken in Singapore (Hii et al., 2009), Puerto Rico (Johansson et al., 2009) and Costa Rica (Fuller et al., 2009). Starting with the study in Singapore (Hii et al., 2009) the analysis reveals a comprehensive approach of the biological evolution of the *Aedes* species depending on weather conditions through the implementation of a time series Poisson regression with over dispersed term. Modelling was based on smooth cubic spline functions capturing the shapes of climatic predictors of mean temperature and cumulative rainfall weekly aggregated. Furthermore an autoregressive term for:

incubation, infectious and survival time estimations was adjusted on the time series. As a result, once the average temperature exceeded 27.5°C the effect on the incidence rate showed a linear growth after 5 to 16 weeks. Unlike the effect on the incidence with temperatures below 27.8°C the delay time was found significant after 17 to 20 weeks. Regarding the impact of rainfall, the investigation undertaken in Singapore, revealed an increased linear pattern when rainfall is below 75 mm but a sharp decline when rainfall exceed the 75mm level. As a result a quadratic form is disclosed as U shaped.

Another example of a non-linear association in weather conditions associated with the severity of the disease was found in Puerto Rico using reported cases over a period of 20 years at municipality level (Johansson et al., 2009). A distributed lag approach with lagged estimation for climatic variables of temperature and rainfall using a hierarchical framework was performed in this investigation. Those findings revealed a significant range between 0 to 2 months based on maximum and minimum temperature and a significant time lag of 1 to 2 months regarding the rainfall effects. The inclusion of population density and poverty levels at the municipal level, provided a more integrated analysis in Puerto Rico.

Furthermore, a study performed in Costa Rica (Fuller et al., 2009) adopted a Cross Correlation Analysis to assess the impact of the El Nino index, Sea Surface Temperature (SST) on dengue cases, by assuming a sinusoid partnership association regarding the climatic index. These findings revealed a strong association with the El Nino3.4 index after 11 weeks. However a lower association but still significant was found regarding the El Nino 2, El Nino 3 and El Nino 4 after 5, 10 and 17 weeks respectively of the occurrence.

In the same vein, a study conducted in Australia (Hu et al., 2010) implemented a time series approach using a seasonal Autoregressive Integrated Moving Average (SARIMA) amongst dengue cases and SOI index from 1993 to 2005 with an Early Warning predictor using forecast of the ENSO. These findings revealed a significant

negative association regarding the SOI and dengue cases with a range of delay time from 3 to 12 months.

Climatic patterns of the El Nino Southern Oscillation phenomenon identified in similar geographical location, suggests that El Nino3.4, could be the best indicator for the ENSO phenomenon on the north-western part of Venezuela, assuming a delay time of 6 months before the appearance of the outbreaks. This statement can be validated by a previous study performed in Brazil (Lowe et al., 2011).

2.5.4 Socioeconomic factors

Aetiological studies of dengue cases associated with climatic factors have been an invaluable assistance for monitoring and preventing the transmission of dengue fever worldwide (Johansson et al., 2009; Hii et al., 2009; Hopp and Foley, 2001; Patz et al., 1998). However experts state that other non-climatic factors need to be included within the analysis to provide a more integrated surveillance system of the disease (Fuller et al., 2009; Hu et al., 2010; Patz et al., 1998; Hopp and Foley, 2001).

In this regard a selective set of studies analysed in this investigation, identified as relevant socio-economic factors: the population density, unplanned urbanization programs, ineffective control programs, poverty, absence of sanitary services and ineffective health infrastructure. All of these have been summarized in Table 2.7.

Worldwide	Factors	Reference
	<ul style="list-style-type: none"> • absence of water supply • unplanned urbanization • ineffective control of disease 	(Guha-Sapir and Schimmer, 2005)
	<ul style="list-style-type: none"> • population density • poverty • poor control of disease • ineffective water supply • human migrations 	(Patz et al., 1998)
Venezuela		
	<ul style="list-style-type: none"> • ineffective control of disease • poor housing conditions • lack of sanitation service • unsuitable health infrastructure 	(Brightmer and Fantato, 1998)
	<ul style="list-style-type: none"> • frequency-water supply • frequency-garbage collection 	(Barrera et al., 2002)
	<ul style="list-style-type: none"> • literacy rate • sewerage service • garbage collection 	(Monsalve et al., 2010)
Brazil		
	<ul style="list-style-type: none"> • urban population • % houses with water supply 	(Lowe et al., 2011)
	<ul style="list-style-type: none"> • garbage collection • water supply • sewerage service • population density • persons per house • education level 	(Mondini and Chiaravalloti-Neto, 2008)
	<ul style="list-style-type: none"> • people living in slums • public water service • garbage collection • literacy rate 	(Ferreira and Schmidt, 2006)
Thailand		
	<ul style="list-style-type: none"> • garbage collection • population density • % slums • % empty houses 	(Thammapalo et al., 2008a)
India		
	<ul style="list-style-type: none"> • frequency-clean water containers • features house • water container covered • mosquito nets • frequency water-supply • solid waste 	(Bohra and Andrianasolo, 2001)

Table 2.7: Other aetiological factors associated with dengue fever in a socioeconomical context

By a meticulous analysis of each one, the study undertaken in India (Bohra and Andrianasolo, 2001), revealed a strong dependence amongst incidence rates and certain cultural practices within the community. This study assumed a linear association amongst the predictors and health outcome. Consequently a Pearsons coefficient was tested to measure such strength. These findings confirm that certain communal habits increase the risk of epidemics based on carelessness in cleaning water containers used for washing and bathing.

The study from India also revealed that much of the population leave the water storage containers uncovered, which become a good breeding site of mosquitoes. This investigation also showed that protection against the insect bite is essential to avoid the transmission of the virus. The most effective actions according to the authors being: spraying repellent against a swarm of mosquitoes, keeping doors and windows closed and the use of nets as preventive measures indoors. Another factor considered as being a trigger for the disease in India is the presence of exposed solid waste in the vicinity of the households, including: bottles, cans, plastic containers and/or old tyres which become perfect breeding sites for mosquitoes. These findings have been confirmed as outstanding factors associated with the spread of the disease in countries like Venezuela (Barrera et al., 2002; Monsalve et al., 2010), Brazil (Ferreira and Schmidt, 2006; Mondini and Chiaravalloti-Neto, 2008) and Thailand (Thammapalo et al., 2008a,b).

Some other investigations conducted in different part of the world spotlight a strong association between absence of public water supply and severity of the disease (Ferreira and Schmidt, 2006; Mondini and Chiaravalloti-Neto, 2008; Lowe et al., 2011; Barrera et al., 2002; Guha-Sapir and Schimmer, 2005; Patz et al., 1998). This association is because the absence of water supply promotes the water storage in artificial containers and consequently breeding sites for *Aedes* species (Caprara et al., 2009). Academics claim that the lack in sanitary services result from an inconsistent plan of urban development (Guha-Sapir and Schimmer, 2005; Patz et al., 1998; Brightmer and Fantato, 1998). As a result more studies are needed to fully understand the association of dengue

transmission with a wide range of socioeconomic factors.

2.5.5 Geographic factors

Epidemiological studies of dengue fever have benefited during the last years from the growing development of satellite imagery (Chang et al., 2009). Visual imaging can easily capture the geographical features of locations at risk. Altitude (metres above sea level) is one of the most important factors to be considered in connection with dengue transmission (Ferreira and Schmidt, 2006; Barrera et al., 2002; Lowe et al., 2011).

Some authors claim that the severity of dengue fever in Venezuela is due to the extensive length of coastline, marked by the lowest altitude levels in conjunction to hot and dry weather conditions (Barrera et al., 2002). In this context some investigations have found a strong association between low altitudes and the spread of dengue fever in different parts of the world (Diaz et al., 2007; Lowe et al., 2011).

It has been stated that biological conditions prevent the *Aedes* species from flying at higher altitude according to some experts (Ferreira and Schmidt, 2006). However, some evidence has recently found cases in high mountainous regions in Asia, Africa and Latin America (Epstein et al., 1998). In particular in Argentina (Porcasi et al., 2012; Gurtler et al., 2009; Carbajo et al., 2012) in which recent outbreaks of dengue fever have alarmed the public health sector.

2.5.6 Demographic factors

Dengue fever is listed as an urban disease because it spreads readily in urban contexts, due to the disproportionate increase of human population (Kyle and Harris, 2008). Numerous studies show evidence of such association (Mondini and Chiaravalloti-Neto, 2008; Barrera et al., 2002; Lowe et al., 2011; Freitas et al., 2003). In the case of Zulia, state this factor might be considered for modelling purposes in further analysis.

2.6 Framework of Early Warning System (EWS) for infectious diseases

Predicting outbreaks of dengue fever has been identified as a main goal to be achieved in the near future according to the WHO (Chan, 2012). The international body has also acknowledged the valuable use of early detection of infectious diseases aimed to effectively managing the interventions on the health sector. And has defined the EWS as a powerful tool to be included on every government agenda.

In this context, the use of satellite remote sensing has greatly benefited the construction of forecasting models based on epidemiological data (Connor et al., 1998) in different parts of the world. Although several studies have contributed to a prompt detection of malaria in different parts of the world (Bouma and Van der Kaay, 1996; Bouma et al., 1997; Connor et al., 1998; Thomson et al., 2004; Abeku et al., 2004) only quite recently that early warning detection has focussed on dengue fever (Lowe et al., 2013; Hii et al., 2012; Descloux et al., 2012; Hu et al., 2010).

The next limitations regarding the construction of EWS need to be overcome in future studies (Rodier et al., 2004):

1. More studies of EWS are required at the National and International level for infectious diseases.
2. The predictive models need to be implemented in different geographical contexts for a more accurate validation.
3. Although a large number of studies have drawn on EWS based on climatic factors, the inclusion of a wide range of non-climatic variables might trigger a more robust performance of the prediction.

4. The EWS needs to lead to the attainment of effective interventions programs for prevention and control of infectious disease.

2.6.1 Protocol for developing Early Warning Systems for infectious diseases

A protocol for developing an EWS of infectious diseases has been proposed by the WHO (Rodier et al., 2004), which has been summarized in Figure 2.4. In accordance with the protocol, before commissioning an EWS of outbreaks in the public health sector, it is essential that there is a reliable surveillance network able to systematically provide epidemiological information in space and time with a high level of accuracy.

Infectious diseases such as: cholera, malaria, dengue fever, yellow fever, west Nile virus have been classed as epidemics so are potential candidates for developing EWS because of their marked seasonality trends (Rodier et al., 2004). In addition these authors claim that understanding the variability of the disease across geographical units over a period of time provides valuable information for the construction of an EWS. As a result areas at risk can be highlighted for a prompt response from the health authorities.

Early detection of epidemics is the key to a good EWS, according to some experts (Connolly, 2005). They suggest at least 5 weeks in advance before an epidemic as ideal time for launching of preventive actions. The aforementioned authors also state that epidemics take place when the cases overcome one standard deviation of the mean over a period for at least two consecutive weeks.

A growing number of statistical and biological models have been widely applied in different parts of the world seeking factors associated with the spread of infectious diseases (Lowe et al., 2011; Yu et al., 2011; Chen et al., 2010; Patz et al., 1998; Barrera et al., 2002; Monsalve et al., 2010). However more accurate approach demand a wide range of variables including: population densities, epidemiological data, entomological data, immunity levels of the patients, poor nutrition, resistance to the pesticide and

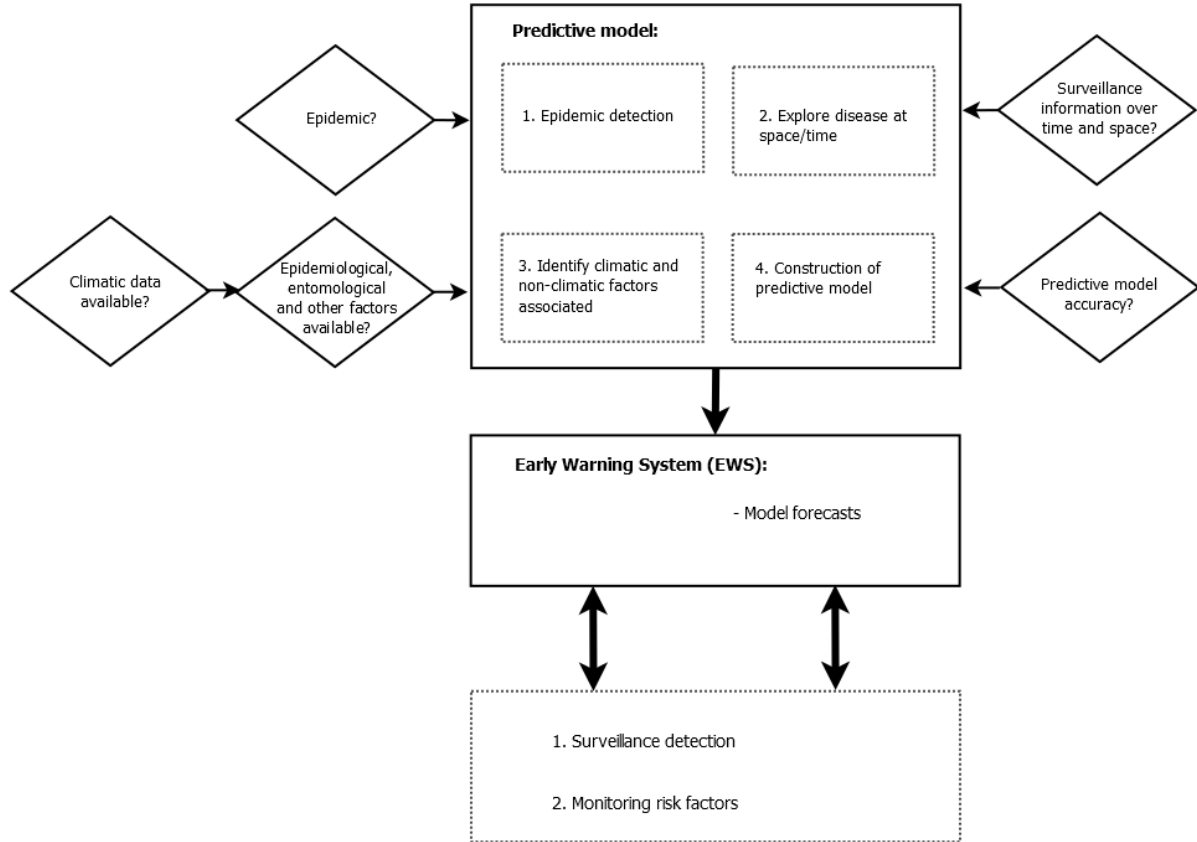


Figure 2.4: Protocol for developing Early Warning System (EWS) of infectious disease environmental and climatic information (Rodier et al., 2004).

Rodier et al. also argue that predictive models do not always lead to an EWS but this step is crucial towards the construction of an EWS once a forecast model is achieved. According to the protocol for developing EWS of infectious disease (see Figure 2.4) some suggestions need to be considered to ensure the effectiveness of an EWS:

1. Fitting the model by excluding effects observed over a long period of time and periodic variation.
2. Implementing the prediction with dataset outside to the proposed model for val-

idation purposes.

3. Quantifying the accuracy of the predictions.
4. Build forecast model based on the prediction of explanatory variables.

In addition, as seen in Figure 2.4 the EWS depends on a systematic input from the surveillance system to ensure quantity and quality of monitored data.

2.6.2 A review of Early Warning System of Dengue fever

A global challenge commissioned by the WHO for the next 7 years is the prevention and control of dengue fever based on information provided by Early Warning Systems validated with highest standards of efficiency (Chan, 2012). In this context, this section discusses the various EWS models of dengue fever developed in different parts of the world stressing advantages and potential improvements.

A study conducted in New Caledonia (Descoux et al., 2012), archipelago located on the Pacific Ocean considered entomological data associated with weather conditions in a non-linear fashion via time series analysis. The accuracy of the prediction was endorsed with 74% of accuracy. However the validation of the model with data outside the model was not performed, becoming a limitation that need to be overcome in further studies. Moreover as the authors described, a potential improvement could be the expansion of the methodology by covering some other geographical contexts.

A recent study conducted in Singapore (Hii et al., 2012) applied a time series Poisson regression climate-based forecast model of dengue cases using spline functions to represent the non-linearity on the covariates (Hii et al., 2012). This approach represents a useful tool, to timely identify epidemics within the region. The forecasting model was built through the implementation of a time series framework based on weekly data from 2000 to 2010 validated with a new set of information covering 2011 to 2012. Furthermore an evaluation of sensitivity of outbreaks was conducted to ensure effectiveness of

the model. According to the authors the forecast version extrapolated the distribution of cases because the last trend of dengue fever largely affects the future. In this respect, many lessons about the forecast modelling in Singapore need to be considered for future studies of EWS in different parts of the world. The weakness of this approach lies in the omission of non-climatic factors, which might contribute to a better understanding of the predictive process (Rodier et al., 2004).

A similar approach was built in Southeast Brazil via a Negative Binomial Generalized Linear Mixed model (GLMM) with a Bayesian framework using WinBUGS tool (Lowe et al., 2013). This study developed an EWS with a set of climatic and non-climatic covariates at spatial and temporal scale from April 2001 to December 2007. In addition the forecasting model of dengue outbreaks was validated with data outside the model (January 2008-December 2009). According to the authors, the hierarchical structure of the Bayesian framework allows the inclusion of unstructured random effects, to account for information which does not belong to the model such as: immunity of patients, serotypes circulating within the region, intervention of control, amongst others. While the structured random effects allows for autocorrelated structures through the implementation of autoregressive (CAR) model. Benefits of forecast modelling using the GLMM via a Bayesian perspective is based on the richness provided by the prior prediction, which lead the posterior distribution of the estimations.

Finally some authors have made suggestions to overcome the boundaries in the construction of an EWS of dengue fever (Lowe et al., 2013; Hii et al., 2012; Racloz et al., 2012; Descloux et al., 2012). The main issues are:

- The information provided by the surveillance system needs to be well geo-referenced with high resolution in time and space. Furthermore the units need to be unified to avoid inconsistency.
- The tendency in the construction of EWS is climatic-based. However this ap-

proach avoids an embodiment of a more robust version. The inclusion of non-climatic factors such as: programs of prevention, serotypes circulating, population size, socioeconomic factors, immunity process need to be included as potential factors for further studies.

- Because historic patterns on dengue cases have a tendency to recur over time, previous cases can be used as information for modelling the posterior prediction via a Bayesian framework.
- Validation of forecasting models needs to be a subject of rigorous assessment including information out-of-model. Flaws in epidemic detection might lead to tragic consequences in the public health and the economic sector.
- Delay times of climatic factors need to be re-scaled within a forecast modelling.
- The EWS needs to be extended to different geographical contexts.

Chapter 3

Exploratory data analysis

3.1 Introduction

In this chapter an exploratory analysis was conducted following guidance (Zuur et al., 2009; Zuur, 2012) from a protocol of data exploration described in detail in chapter 2, section 2.4.

A preliminar analysis regarding the epidemiological data of dengue cases in Zulia state from 2002 to 2008 was performed to highlight the most relevant features on the dynamics of dengue transmission. Moreover the exploration of potential explanatory variables associated with the disease was also conducted.

Although the complexity on dengue transmission has not been fully understood (Gubler, 1998a,b; Guha-Sapir and Schimmer, 2005). A set of explanatory variables was selected based on previous studies undertaken in different parts of the world. They included: demographic (Mondini and Chiaravalloti-Neto, 2008; Barrera et al., 2002); socioeconomic (Diaz et al., 2007; Bohra and Andrianasolo, 2001); geographic (Barrera et al., 2002; Lowe et al., 2011); climatic (Lowe et al., 2011; Chen et al., 2010; Fuller et al., 2009; Kovats, 2000; Githeko et al., 2000) and individual risk factors (Raja and Devi, 2006; Sierra et al., 2007; Kongsomboon et al., 2004; Morrison et al., 1998).

The findings from this chapter will support subsequent analysis at the municipal level in Zulia state to assess the association between explanatory variables and the distribution of the disease over space and time (Lawson, 2009; Hoeff and Boveng, 2007; Zuur et al., 2009; Hii et al., 2009; Johansson et al., 2009; Fuller et al., 2009; Wood, 2006). In summary, the main goal in this chapter is to use the available dataset to pursue an exploratory analysis to test which variables should be taken forward for examination in greater depth through more sophisticated modelling.

3.2 Research objectives

- Undertake an exploratory analysis to investigate patterns of the dynamic of dengue fever in Zulia state and potential factors associated with the disease.
- Investigate collinearity amongst the explanatory variables, which could assist the construction of subsequent models.
- Explore the association between dengue transmission and weather conditions in Zulia state, to justify the use of a more flexible approach without parametric assumptions.

3.3 Research questions

- What are the explanatory factors associated with dengue fever in Zulia that merit testing with more advanced statistical methods in subsequent analyses?
- Under the assumption that the association between climatic factors and dengue transmission is unknown, does the data gathered for Zulia state provide adequate information regarding the real shape of such an association?

3.4 Data exploration of the dynamics of dengue fever in Zulia-state

Zulia state is located in the Northwest part of Venezuela, between 8.20° and 11.79° South-North latitude and between -70.73° and -73.37° East-West longitude. It is divided into 21 municipalities over an area of $50,23 \text{ Km}^2$ surrounding the Lake Maracaibo. This lake is the largest inland body of water in Latin America, covering $12,870 \text{ Km}^2$.

Zulia state is also divided into five main regions (see Figure 3.1). Each of these regions comprise an administrative division characterized by having similar socio-economic development (Alvarez, 2009). These are: Guajira, East Coast, Capital, Perija and South.

Guajira, is located in the north of Zulia and includes the municipalities of: Paez, Mara and Padilla.

East Coast, is located on the east coast of Lake Maracaibo and includes the municipalities of: Santa Rita, Cabimas, Simon Bolivar, Lagunillas, Valmore Rodriguez, Baralt and Miranda.

Capital, includes the capital Maracaibo along with Urdaneta, San Francisco and Los-sada municipalities. It is located in the west coast of Lake Maracaibo

Perija, located in the central area. It is formed by Rosario and Machiques.

South, located on the south of Lake Maracaibo, comprises the municipalities of: Semprun, Catatumbo, Colon, Pulgar and Sucre.

The maps below show the location of the municipalities belonging to each region.

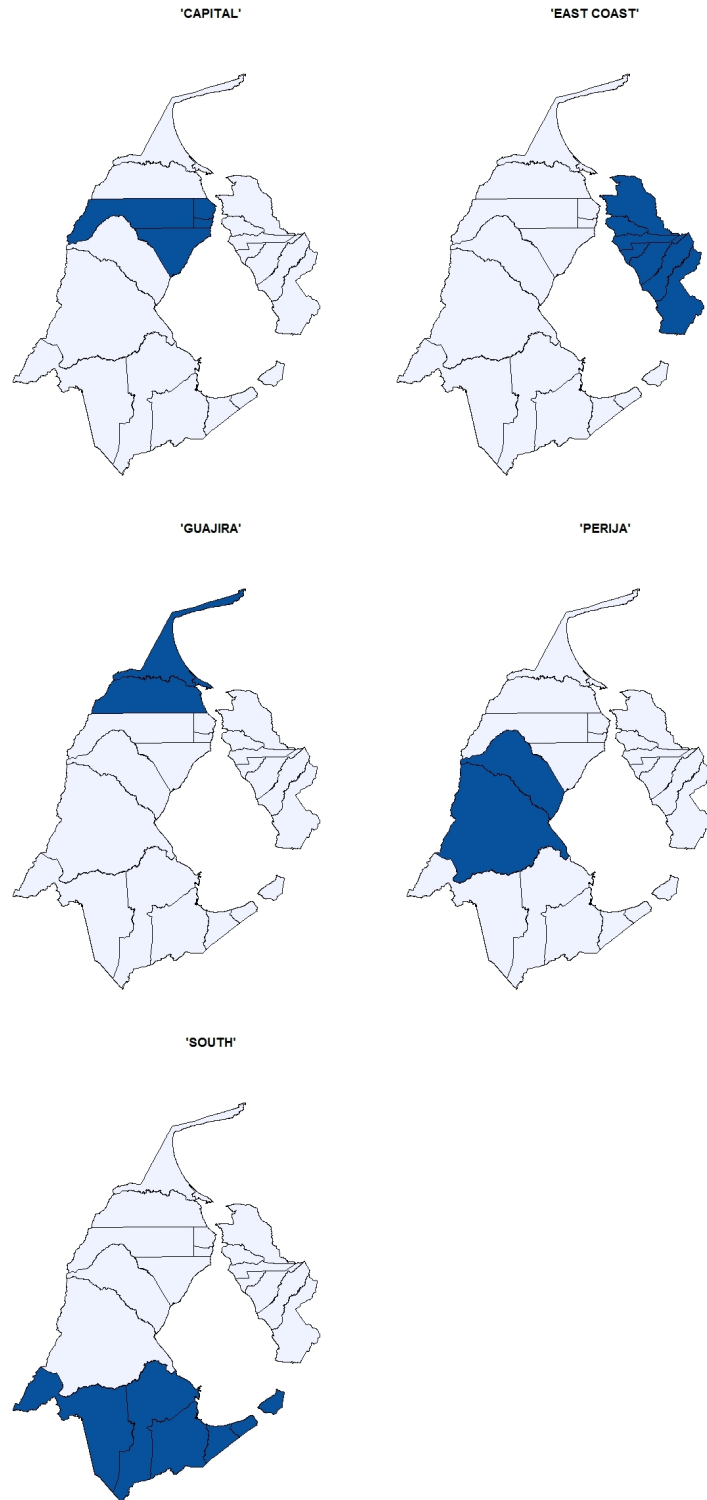


Figure 3.1: The five main administrative regions which comprise Zulia state

The data was arranged in many different ways throughout the exploratory analysis. In the first instance a study was based on epidemiological records of hospital admissions weekly aggregated from January 2002 to December 2008. The reason for such an approach was because the data was initially provided at this scale by the local agency of the Ministry of Health in Venezuela (see section 2.3.2). In addition annual population size was interpolated to fill weekly and monthly gaps for the missing data. As a result epidemiological data in the whole of Zulia was initially aggregated per week covering the period from January 2002 to December 2008. So the incidence rates were also weekly estimated through the ratio between dengue cases and the corresponding population size.

Figure 3.2 shows a time series plot weekly aggregated from January 2002 to December 2008 of dengue incidence in the whole of Zulia state. These findings revealed a temporal variation on the incidence rate reaching a peak of 3.48 per 10,000 inhabitants in 2007. In average a number of 378 cases per week during that year. Large values on the incidence of the disease were also seen during the first 10 weeks in 2002, 2007 and 2008. But very low values in 2003, 2004 and 2005 across the 52 weeks. Moreover a sharp severity of the disease was observed at the beginning and the end of 2007. Likewise a rising trend was observed in 2006 by the end of the 52 weeks.

Although some research conducted in Venezuela has noted the severity of the disease during 2006 and 2007 in Zulia state (Oletta, 2006), the underlying causes have not been subject to systematic investigation in the Northwest part of Venezuela.

However the outbreak of the disease visible early and late in some years might follow a seasonal explanation associated with the climatic factors (Nakhapakorn and Tripathi, 2005). Moreover the inter-annual fluctuation of the disease especially during 2006 and 2007 might be associated with the impact of the global climatic phenomenon of El Nino Southern Oscillation (ENSO) (Lowe et al., 2011; Yu et al., 2011; Gagnon et al., 2001). Hence these matters warrant further probing based on the dataset from

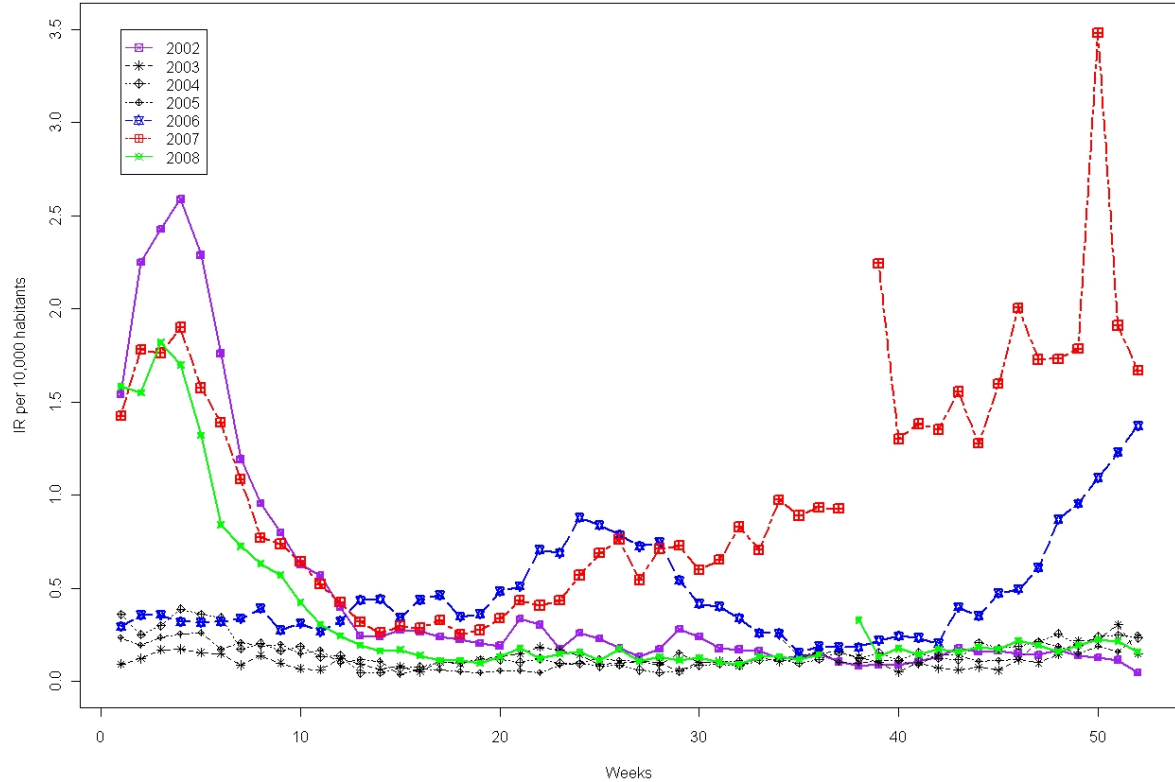


Figure 3.2: Distribution of dengue incidence weekly aggregated per 10,000 inhabitants in Zulia from 2002-2008

Zulia state.

Following a new arrangement of the dataset, a visualization of the crude incidence rate in Zulia state at the municipality level annually aggregated from 2002 until 2008 (see Figure 3.3) was performed. It portrays the spatial and temporal variation of the disease over the underlying period. Although this depiction is quite crude. There is a considerable body of evidence that ensures drawing up geographical maps is a useful starting point to identify potential factors associated with dengue fever (Freitas et al., 2003; Arboleda and Peterson, 2009; Bohra and Andrianasolo, 2001; Lozano et al., 2008; Barrera et al., 2000). The spatio-temporal distribution of dengue incidence in Zulia state aggregated at municipality level from 2002 to 2008 was plotted using a free Geo-

graphic Information System (GIS) software (Eisen and Lozano-Fuentes, 2009; Morrison et al., 1998) called Quantum GIS (GIS, 2009).

The findings in Figure 3.3 display a wide variation of the incidence in Zulia during the years 2002, 2006 and 2007 as a result of increased cases, across the spatial units. Unlike the pattern for 2002, the years 2003, 2004 and 2005 seem to be more homogeneous with reduced level of incidence amongst municipalities. The map for 2007, shows that the disease was highly dynamic at municipal level. However the reasons for such variation are still poorly understood.

In addition, Figure 3.3 shows a critical situation regarding dengue transmission, mainly located in the north coastal area of Zulia state close to the maritime channel that connects Lake of Maracaibo with the Gulf of Venezuela during 2002, 2006 and 2007. Furthermore a high incidence rate was clearly noticeable in the southwest part of the state during 2007, needing further investigation.

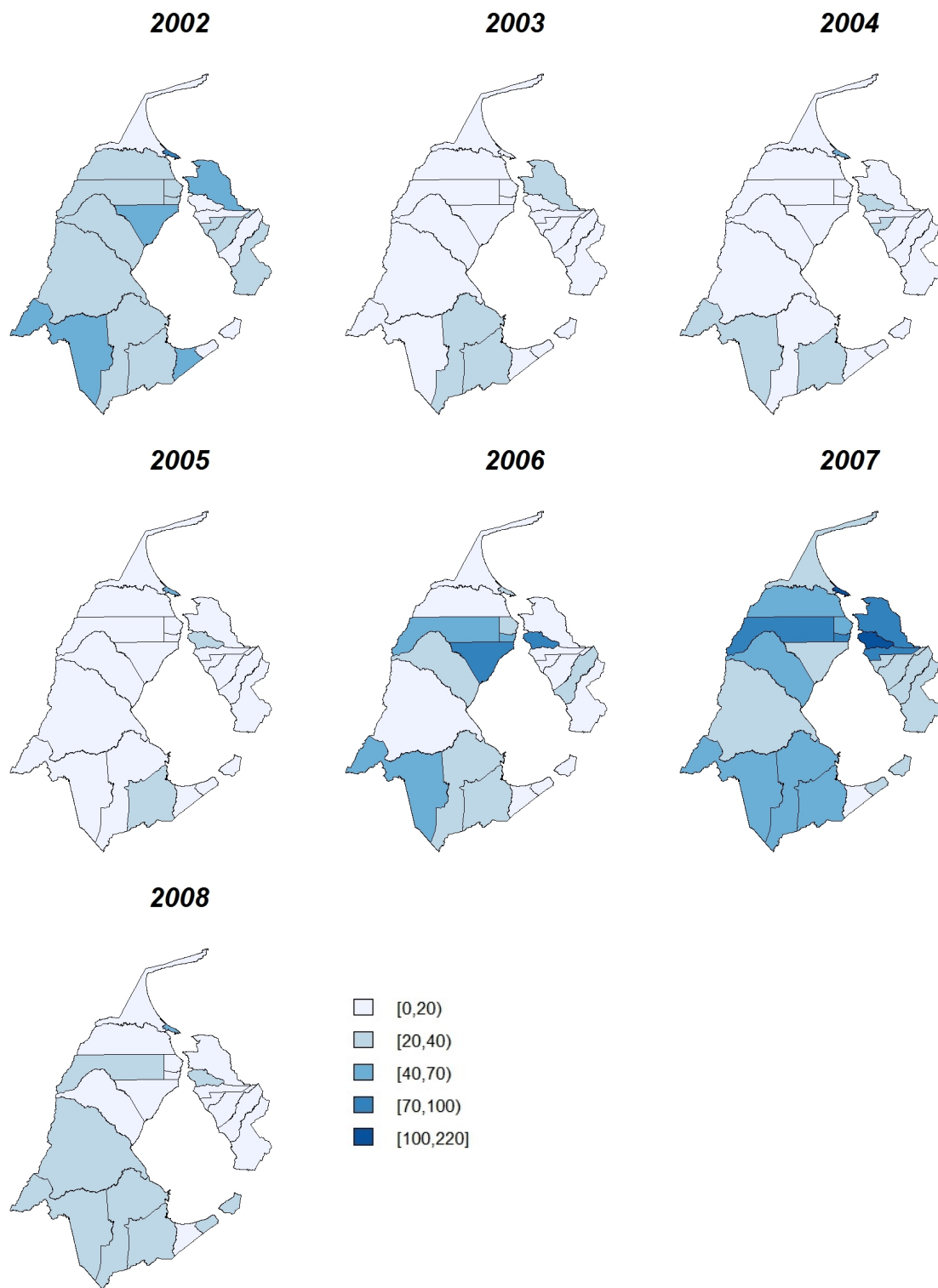


Figure 3.3: Mapping of crude incidence rates of dengue per 10,000 in Zulia state from 2002 to 2008

3.4.1 Data exploration using individual factors

Researchers who have focused on risk factors of dengue fever in different parts of the world have found differences regarding the occurrence of the disease within groups of age, gender and race (Raja and Devi, 2006; Sierra et al., 2007; Kongsomboon et al., 2004; Morrison et al., 1998).

In the present study, the dengue cases were grouped according to age groups and gender within the whole of Zulia state, from 2002 to 2008. Unfortunately a similar analysis could not be carried out at the municipality level owing to poor collection of dengue records aggregated at age group and sex by those health authorities responsible for surveillance. In this regard the dataset under analysis comprises dengue cases in Zulia state over the 364 weeks covering 2002 until 2008, broken down into five age groups, namely: 0 to 4 years, 5 to 14 years, 15 to 24 years, 25 to 44 years and over 45 years old.

Figure 3.4, shows that the age group at risk of the disease are children between 5 and 14 years and the second and third most affected group were those in the 15 to 24 years and 0 to 4 years groups respectively. The statement that those in middle childhood are most at risk in Zulia state, confirms similar patterns observed in Thailand and Florida, in which the highest reported cases are within the population under 15 years old (Kongsomboon et al., 2004; Morrison et al., 1998).

Some researchers have suggested that the severity of the disease in the younger population may be associated with biological factors affecting the genetic sensitivity of such at risk population (Kongsomboon et al., 2004). Others argue that it might be associated to environmental conditions or tendency to group play activities, which might facilitate the spread of the disease (Morrison et al., 1998).

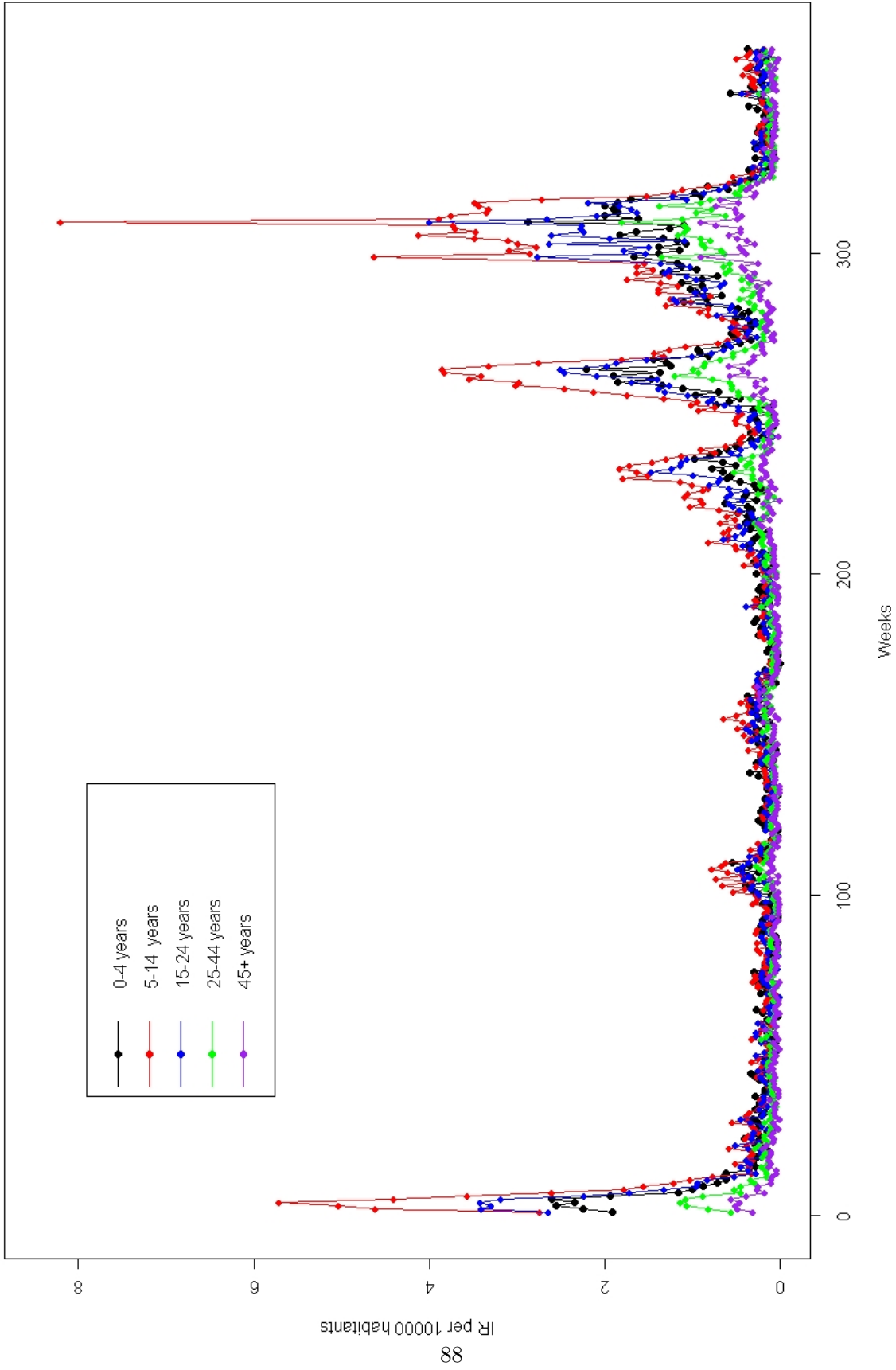


Figure 3.4: Observed incidence rate of dengue fever weekly aggregated by age group in Zulia from 2002 to 2008

On the other hand, the incidence rates per 10,000 inhabitants grouped by gender in Zulia state was displayed in Figure 3.5. A similar pattern of the incidence rate was seen in both male and female groups. These findings are consistent with studies undertaken in some South American countries (Guha-Sapir and Schimmer, 2005) in which dengue cases amongst male and female groups did not show a significant variation. Although a slight preponderance in 2007 within the male population in Zulia is a finding that replicates outcomes obtained in Malaysia (Raja and Devi, 2006). The particular case in Zulia state might be subordinated to a clear outlier event.

Subsequent studies point to the use of Generalized Linear models (GLM) with Poisson distributed data (see section 2.4.2) as the most optimal approach when dealing with dengue cases. This approach might accomplish the inclusion of gender and groups of age as categorical factors.

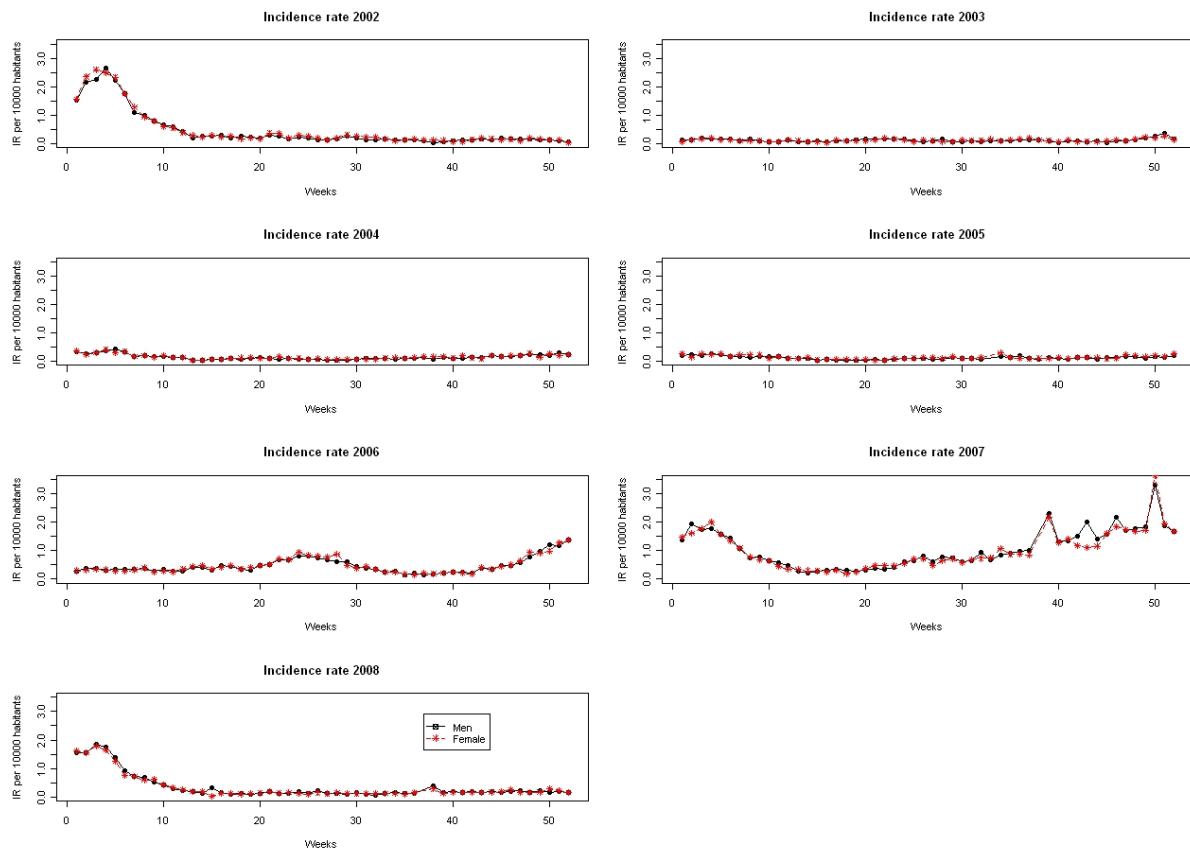


Figure 3.5: Observed incidence rates of dengue fever weekly aggregated by gender in the whole of Zulia from 2002 to 2008

3.4.2 Data exploration with geographic factors

The subsequent exploratory analysis aims to visualize potential variation on count cases of dengue across geographical units. The need to explore similarities and differences of patterns across the available geographics units, led the construction of the following approach at regional level. The following plots show the evolution of the disease in Zulia state nested by the five main regions that comprise Zulia state: Guajira, East Coast, Capital, Perija and South. This approach was based on epidemiological data monthly aggregated from 2002 to 2008 enabling the exploration of potential causality attached to geographic locations.

As a result the Guajira region located on the north part of Zulia, showed the most critical trend of the disease (see Figure 3.6). It was seen that incident rates were widely heterogeneous across the municipalities that belong to Guajira (Mara, Padilla and Paez). In this context, Padilla reached a peak of 51.61 per 10,000 inhabitants during 2007, whilst in the municipalities of Mara and Paez the levels were lower than 10 during the same year. It was seen a marked variation on health outcomes across municipalities pertaining to the same region. This pattern might lead us to believe that region or municipality could be considered as a random effect when fitting a more advanced methodologies (Zuur et al., 2009; Mullen and Birkeland, 2008; Gschlobl and Czado, 2008).

Although the East Coast was the second region in relation to the severity of the disease. A slight variation on trends was also noted regarding incidence rate across the municipalities that comprise this region: Santa Rita, Cabimas, Simon Bolivar, Lagunillas, Valmore Rodriguez, Baralt and Miranda (see Figure 3.7). In addition a peak was reached at 27.32 during 2007 and Santa Rita being the municipality at risk in terms of dengue transmission.

Despite previous trends from Guajira and East Coast regions, the evolution of incidence rate within the Capital region, shows some variation patterns over the mu-

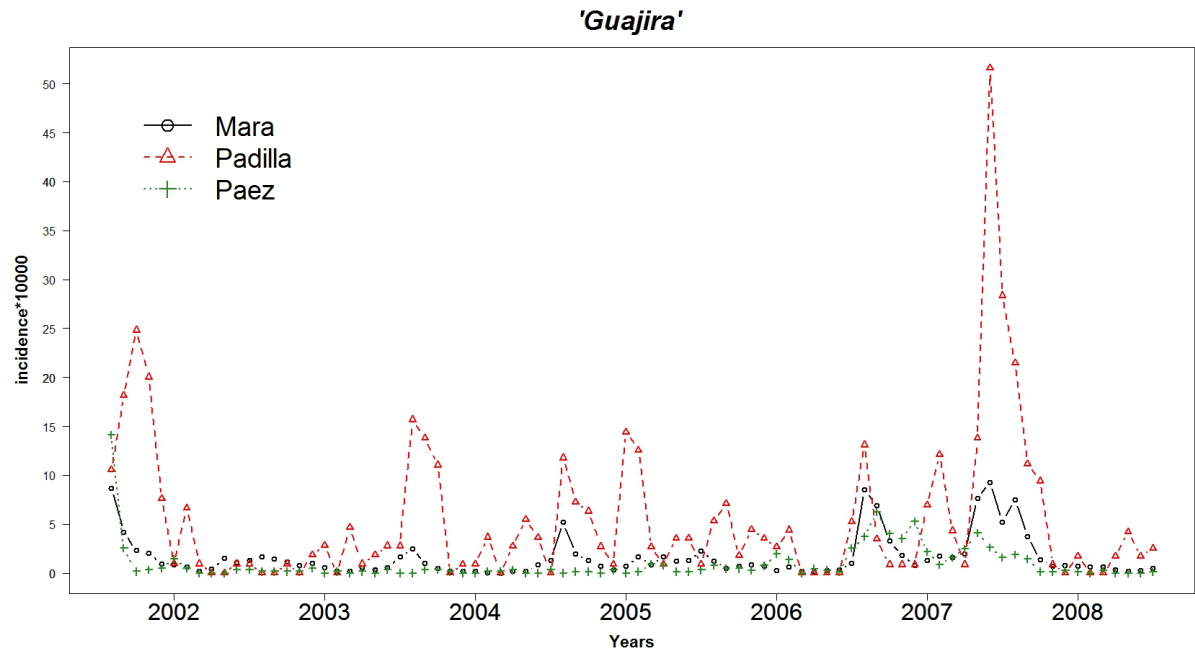


Figure 3.6: Evolution of incidence rate in the Guajira region, monthly aggregated from 2002 to 2008

municipalities within this region: Maracaibo, Urdaneta, San Francisco and Lossada (see Figure 3.8). Urdaneta being the most affected municipality with an incidence of 25.3 per 10,000 inhabitants in 2006.

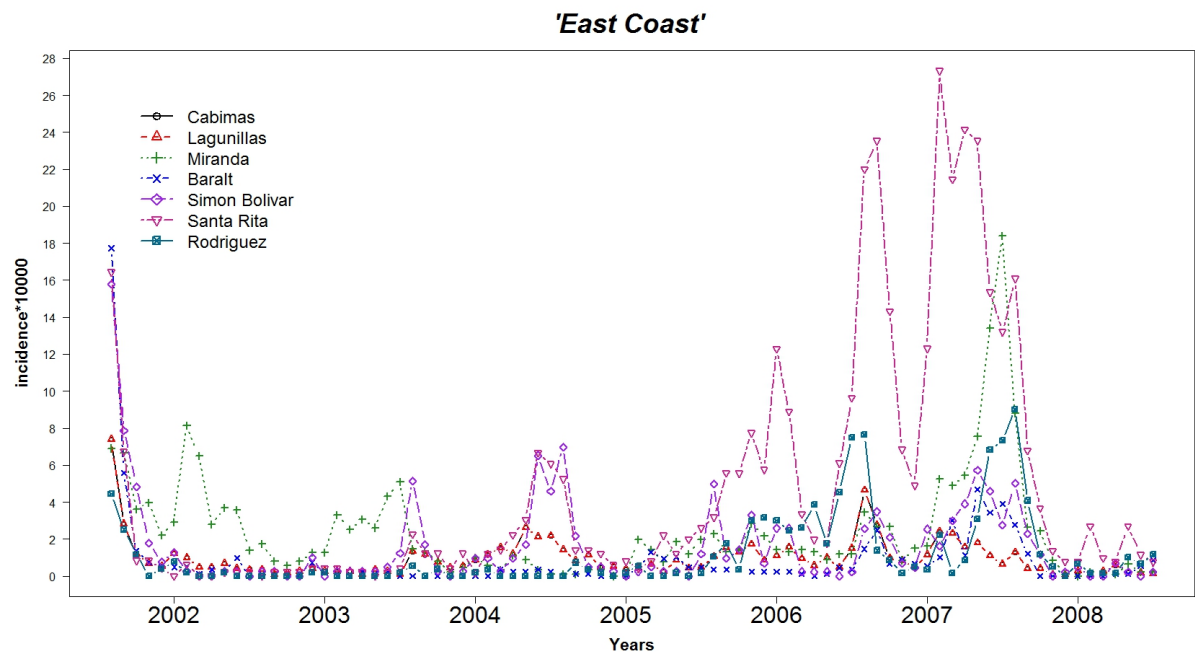


Figure 3.7: Evolution of incidence monthly aggregated in East Coast region from 2002 to 2008

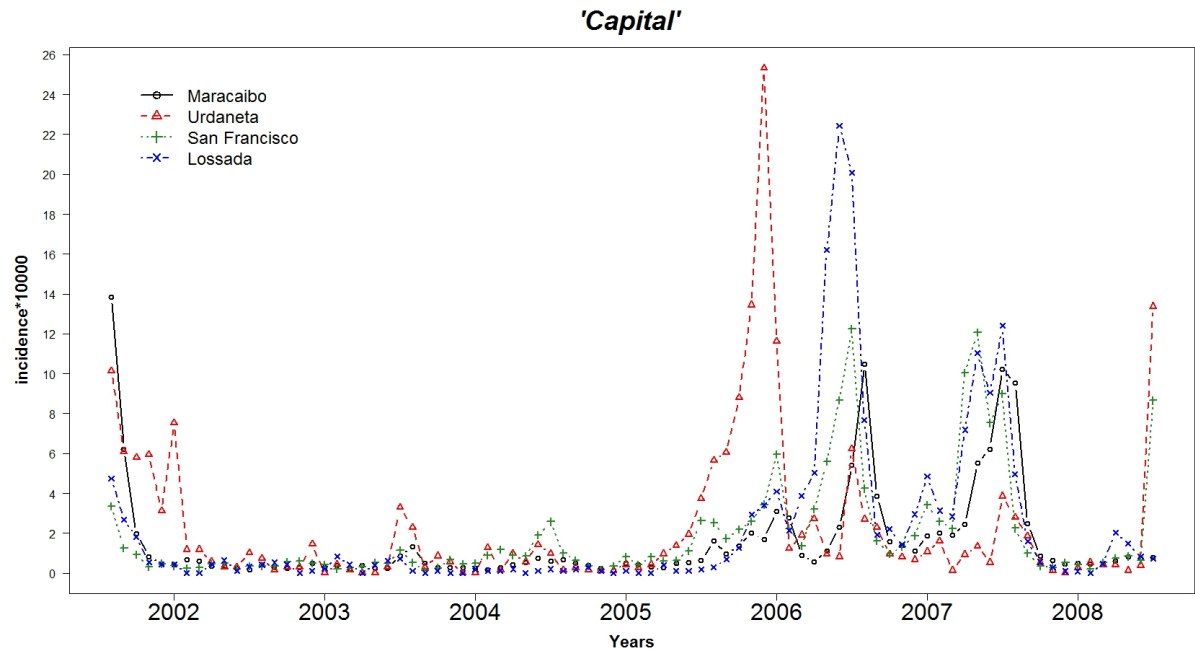


Figure 3.8: Evolution of incidence monthly aggregated in Capital region from 2002 to 2008

However the trends in Perija region display a quite similar pattern on the incidence across the municipalities, as seen in Figure 3.9. Rosario being the municipality at risk with the highest incidence, reached at 22.00 in 2002.

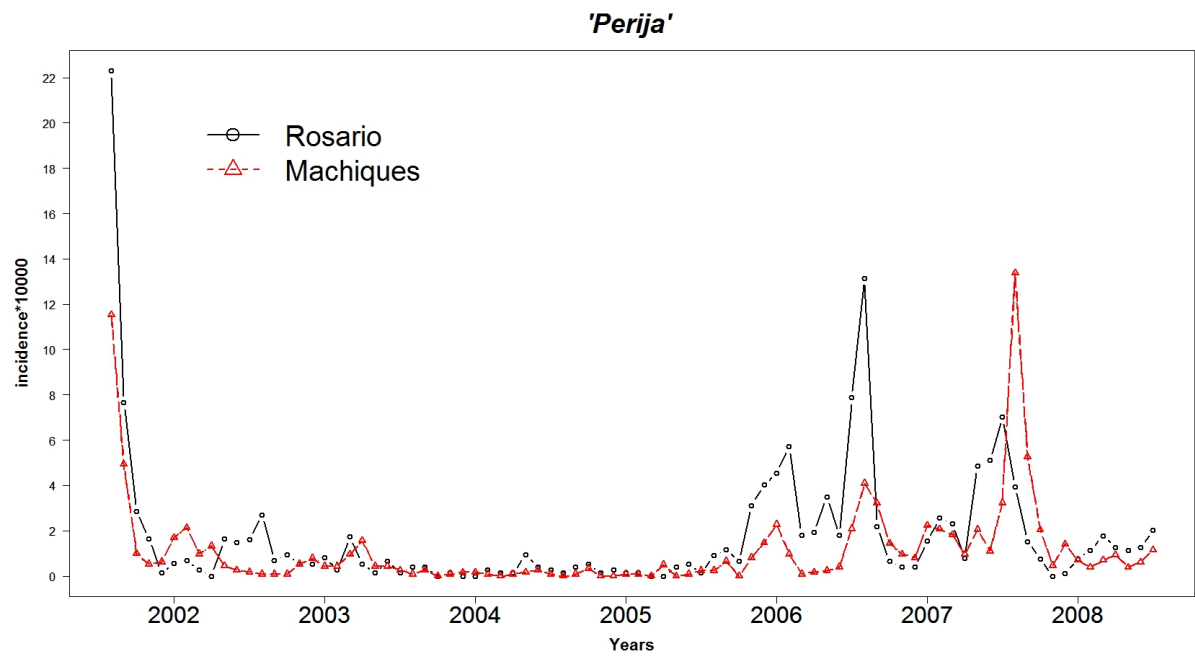


Figure 3.9: Evolution of incidence monthly aggregated in Perija region from 2002 to 2008

And finally another fully unsteady pattern was found in the South region of Zulia state (see Figure 3.10). Although the severity of the disease does not seem alarming. The year 2007 remains as the most critical year. These findings might suggest the possibility to explore in more detail the spatial dependencies amongst municipalities. To this end a subsequent Spatio-Temporal modelling with data from Zulia state could be implement using municipality as a random effect (Wood, 2006; Zuur et al., 2009; Zuur, 2012) to explore the variability across those entities.

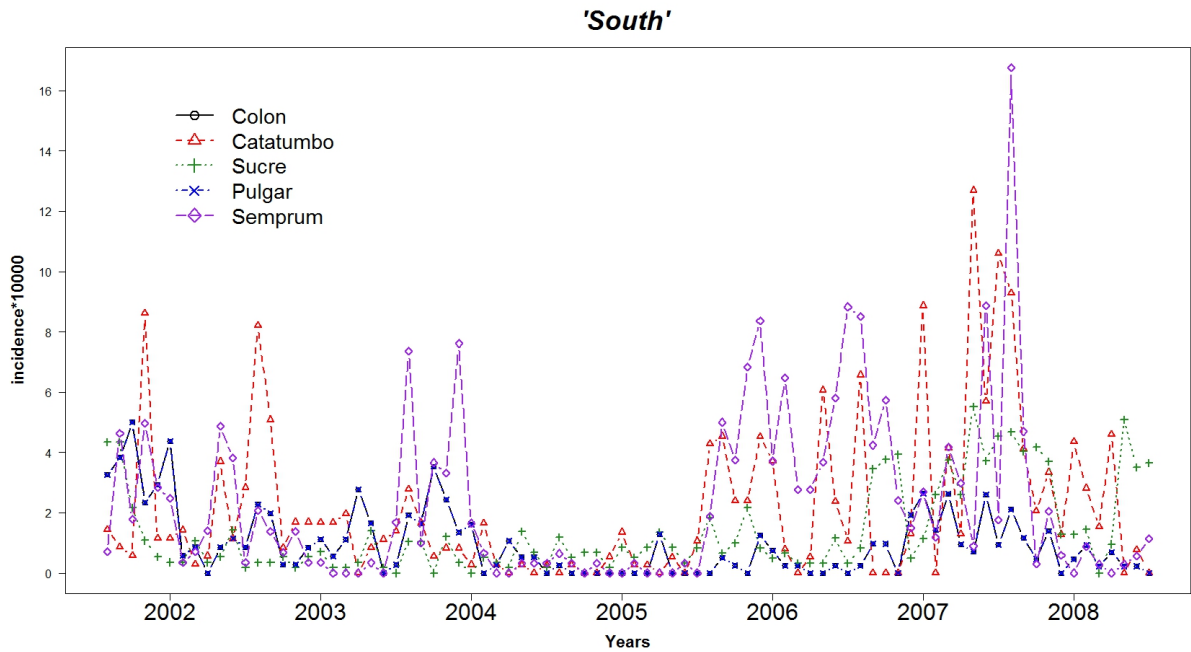


Figure 3.10: Evolution of incidence monthly aggregated in South region from 2002 to 2008

Another relevant geographic factor that needs to be analyzed as a potential explanatory variable associated with dengue transmission in Zulia state is altitude (metres above sea level). Some authors have stated the vulnerability of Venezuela regarding the severity of dengue fever owing to the large kilometres of coast. This feature is marked by low heights and consequently dry and hot conditions (Barrera et al., 2002).

In this context some investigations have found a strong association with lower alti-

tude levels and the spread of dengue fever in different parts of the world (Diaz et al., 2007; Lowe et al., 2011). Biological conditions prevent the *Aedes* species flying over high lands according to some experts (Ferreira and Schmidt, 2006). However, some evidence has recently found cases in regions located on high mountains (Epstein et al., 1998). In particular in countries of the South American cone (Carbajo et al., 2012; Gurtler et al., 2009). Although it is a very complex issue that needs more in depth research, some explanations have been pointed to global warming phenomenon (Wu et al., 2009; Lowe et al., 2011).

To this end a visual representation of average altitude across the 21 municipalities which comprise Zulia state has been displayed in Figure 3.11. It is seen that values fluctuate from 0 to 3,000 metres high. The municipalities located along the coastal line from north to south are distinguished by plain landscape. Conversely the highest elevations are located in the west central side of Zulia state, surrounding the Sierra de Perija. These findings might suggest a potential association regarding low elevations and the severity of the disease in Zulia state, particularly notable in: Padilla, Mara, Paez and Santa Rita located on the north coastline.

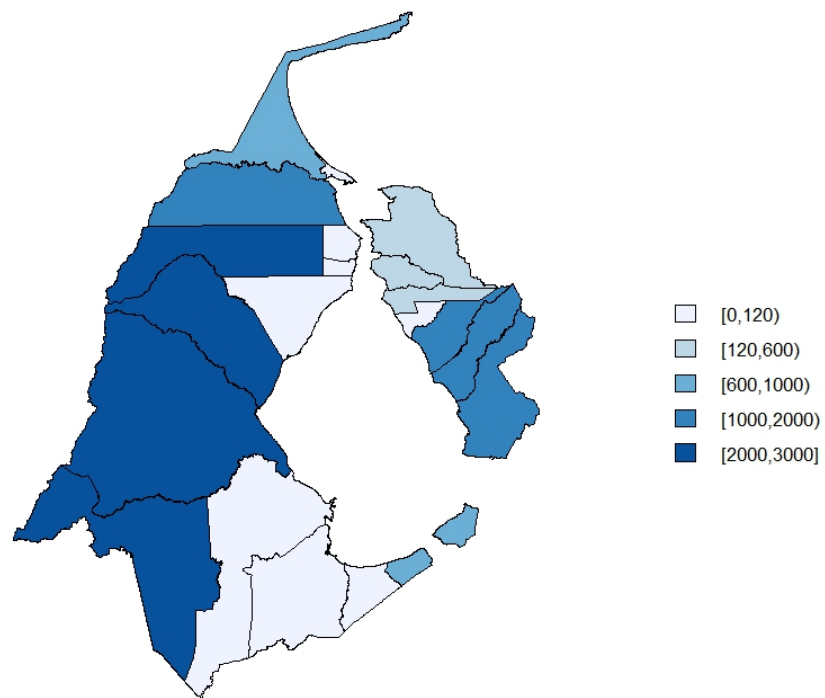


Figure 3.11: Spatial distribution of average altitude across municipalities

Additionally a scatterplot amongst altitude levels and incident rates of dengue fever per 10,000 inhabitants is shown in Figure 3.12. These findings might suggest that lowest levels of altitude affect the incidence rate of the disease in Zulia state. In addition the Pearson coefficient was estimated in 0.25 leading to us considering this variable for modelling of dengue fever in Zulia state.

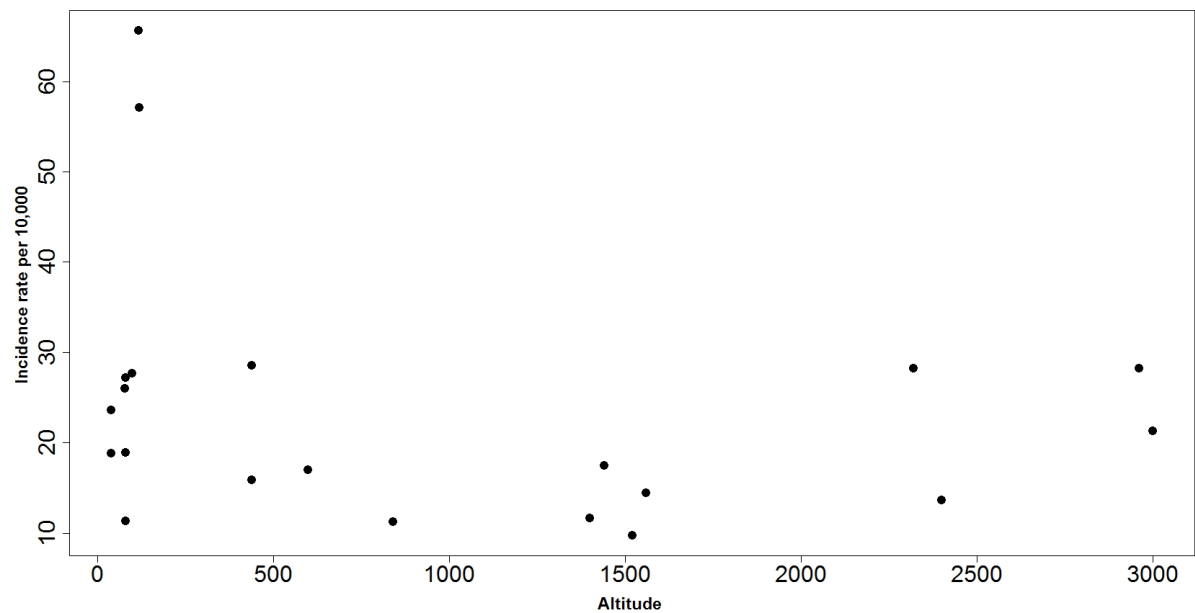


Figure 3.12: Scatterplot of Average altitude versus incidence rates of dengue fever at municipality level

Another variable that merit the attention regarding dengue transmission in the geographical context, is the area at municipality level. To this end, Figure 3.13 illustrates, not only the geographical location of the municipalities but also the territory coverage, expressed in km^2 , which could account for a potential association to overcrowding demographic.

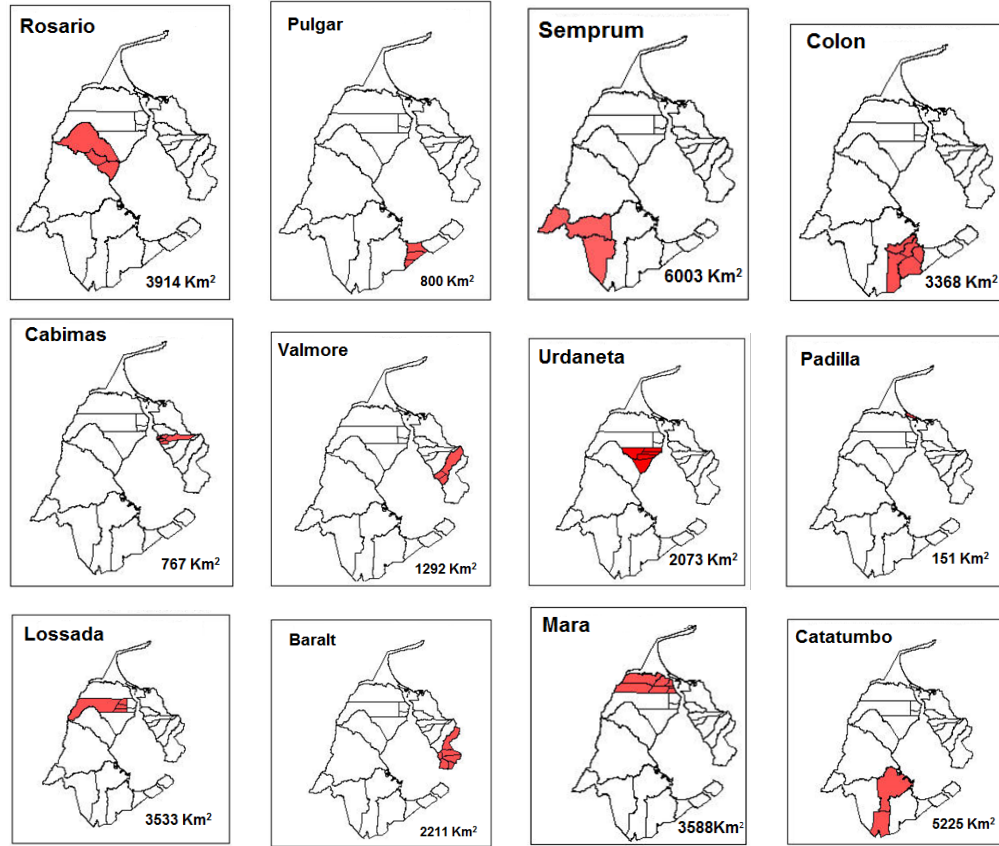


Figure 3.13: Geographical locations of municipalities with territory coverage, expressed in km^2 : (1) Rosario, (2) Pulgar, (3) Semprum, (4) Colon, (5) Cabimas, (6) Valmore, (7) Urdaneta, (8) Padilla, (9) Lossada, (10) Baralt, (11) Mara and (12) Catatumbo in Zulia state

Likewise, Figure 3.14 illustrates the remaining municipalities of Zulia state.

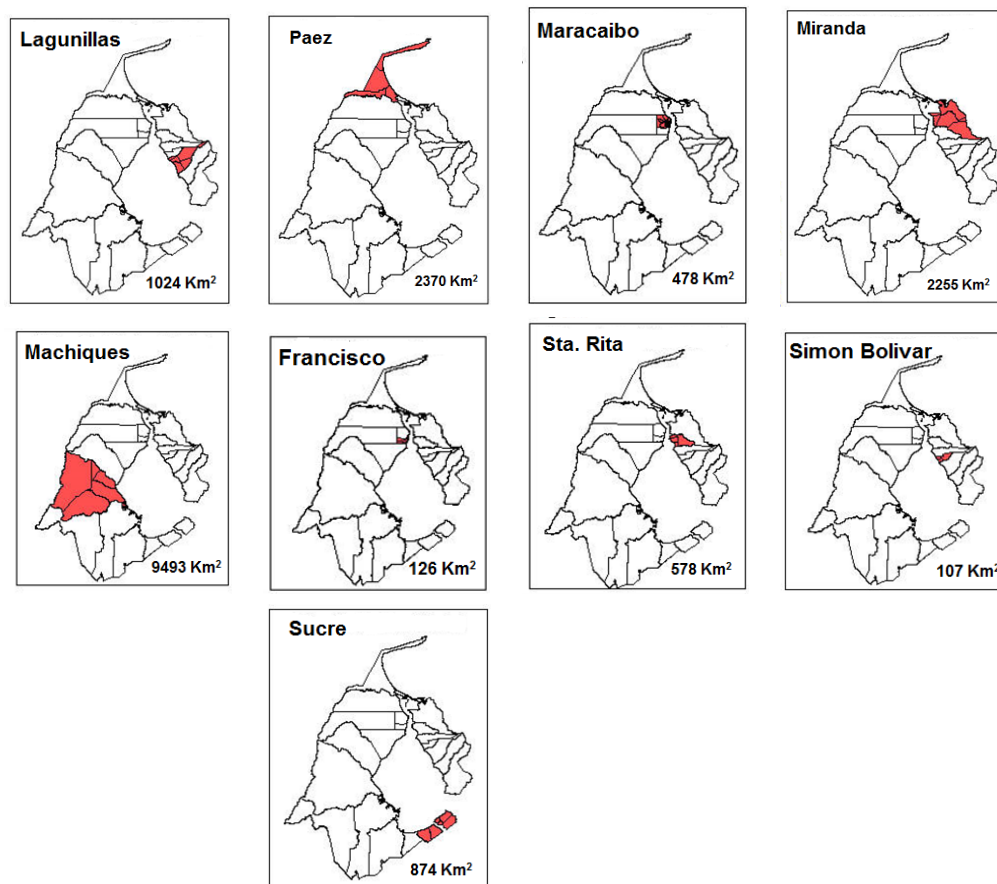


Figure 3.14: Geographical locations of municipalities with territory coverage, expressed in km^2 : (13) Lagunillas, (14) Paez, (15) Maracaibo, (16) Miranda, (17) Machiques, (18) Francisco, (19) Sta. Rita, (20) Simon Bolivar and (21) Sucre in Zulia state

As a result the smaller jurisdictions of: Simon Bolivar, Francisco, Padilla, Maracaibo, Sta. Rita, Cabimas, Pulgar and Sucre seems to have the highest levels of incidence rates. However this statement need to be confirmed in further analysis.

3.4.3 Data exploration using socioeconomic factors

A set of georeferenced data in socio economical context was also examined to pursue an epidemiological study of dengue fever in Zulia state. The information used in this study was based on the National Census 2001 undertaken in Venezuela (see Table 2.2 in chapter 2). As stated in chapter 2, the latest National Census was carried out in 2011. However the outcomes were not published by the time this research was developed.

The selection of variables included within the exploratory analysis in Zulia state, was based on previous studies undertaken in different parts of the world (see section 2.5.3 in chapter 2). They are listed below:

- Percentage of households with access to the piped water supply network
- Percentage of households that receive garbage collection service
- Percentage of unoccupied households
- Percentage of households with domestic air conditioning equipment
- Percentage of households including a bathroom without affordable shower
- Percentage of households that receive electricity service

At first instance, Figure 3.15, shows a pronounced heterogeneity in terms of supplying water pipe network across the 21 municipalities which comprise Zulia state. This pattern might explain the spread of the disease at a spatial scale in Zulia. It is seen that Padilla, Mara and Lossada located on the north of Zulia state in conjunction to Semprun, Catatumbo and Francisco Pulgar on the south are the most neglected jurisdictions on this public service delivery. Although the causes of this situation go beyond the analysis of this thesis. This finding might lead to a potential association with dengue transmission (Mondini and Chiaravalloti-Neto, 2008; Barrera et al., 2002; Guha-Sapir and Schimmer, 2005).

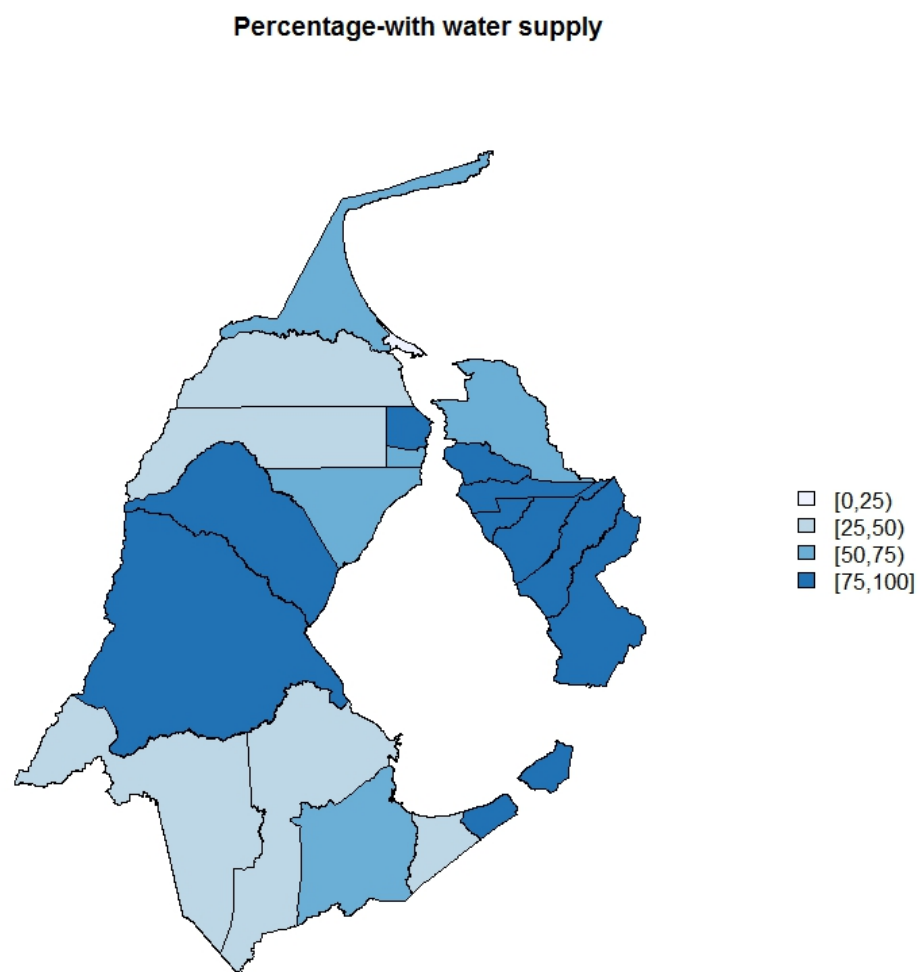


Figure 3.15: Spatial distribution on percentage of households with access to the pipe water supply network

In addition, Figure 3.16, displays a scatterplot between the percentage of households with access to the pipe water supply network and the incidence rates of dengue fever at municipality level. This plot reveals a potential association with dengue transmission due to a Pearson's coefficient estimated at 0.47.

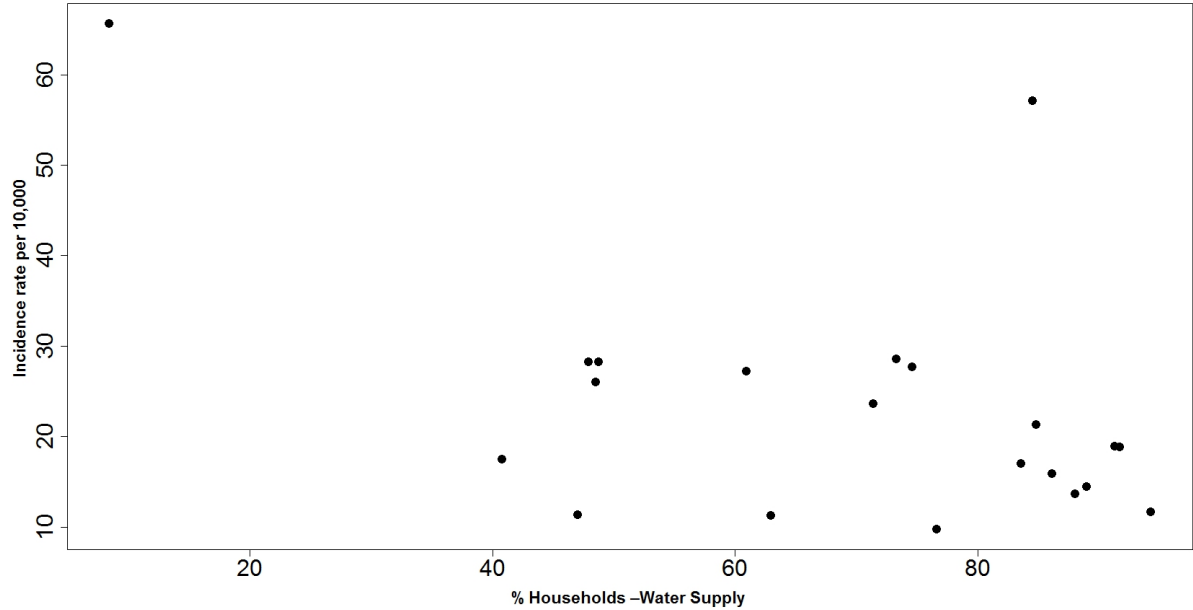


Figure 3.16: Scatterplot of percentage of households with access to water pipe supply network versus the incidence rates of dengue fever aggregated at municipality level

The remaining scatterplots associated with incidence rates are illustrated in A.1-Appendix A. In addition a set of multipanel scatterplots with the aforementioned socioeconomic variables per municipality were also displayed in A.2-Appendix A. But even more, the Pearson's Coefficients were estimated with the dataset in the whole of Zulia amongst the socio economic variables. This study was made to avoid multicollinearity phenomenon in subsequent regression analyses (see A.3-Appendix C). Due to a strong association between the socio economic variables, it is suggested the inclusion of percentage of households with access to the pipe water supply network as the most relevant variable to explain dengue transmission in Zulia state.

3.4.4 Data exploration using demographic factors

Some investigations state that dengue fever expands readily in urban areas due to disproportionate increase of the human population (Kyle and Harris, 2008; Lowe et al., 2011). To this end, it was convenient to list the population size, annually aggregated at municipality level in Zulia state. As it was stated in chapter 2 (section 2.3.2) this information was provided by the Ministry of Health, via the local agency in Zulia (see Table 3.1). These outcomes reveal that Maracaibo has the highest concentration of inhabitants as it is the capital of Zulia state. Followed by San Francisco, Cabimas and Lagunillas, as they are industrial centres of oil production. This information is essential to pursue epidemiological studies. However interpolation processes have been performed to cover the missing data weekly and monthly.

Although most of the studies regarding dengue transmission have implemented the inclusion of population density for modelling the evolution of dengue fever (Barrera et al., 2002; Lowe et al., 2011; Freitas et al., 2003). The inclusion of population size in Zulia state for modelling purposes might be justify because the movement of people across municipalities is still poorly understood, in addition the lack of information in this matter limited its analysis.

Municipality	2002	2003	2004	2005	2006	2007	2008
Maracaibo	1,316,404	1,339,146	1,361,861	1,384,417	1,407,119	1,429,783	1,452,384
San Francisco	378,949	384,799	390,619	396,373	402,147	407,890	413,591
Cabimas	242,939	247,980	253,050	258,120	263,249	268,404	273,580
Lagunillas	186,665	191,673	196,760	201,895	207,130	212,439	217,823
Mara	178,003	183,799	189,729	195,764	201,955	208,281	214,743
Colon	118,782	121,771	124,799	127,850	130,954	134,094	137,272
Machiques	111,084	113,680	116,302	118,937	121,612	124,311	127,034
J.E.Lossada	92,558	94,808	97,087	99,379	101,709	104,062	106,440
Miranda	85,433	86,903	88,372	89,831	91,299	92,764	94,225
Baralt	80,616	82,187	83,763	85,335	86,923	88,516	90,109
Rosario	73,098	74,022	74,937	75,832	76,726	77,607	78,476
Urdaneta	67,029	68,378	69,735	71,089	72,460	73,834	75,213
Paez	63,016	63,941	64,859	65,764	66,671	67,571	68,463
Sucre	55,266	56,371	57,481	58,590	59,710	60,835	61,962
V.Rodriguez	51,554	52,730	53,920	55,113	56,322	57,543	58,772
Sta. Rita	47,453	48,152	48,847	49,531	50,218	50,901	51,575
Simon Bolivar	39,312	40,061	40,814	41,562	42,318	43,077	43,835
Catatumbo	34,653	35,331	36,012	36,693	37,379	38,069	38,758
F.Pulgar	33,772	35,123	36,518	37,951	39,435	40,965	42,542
J.M.Semprum	27,980	28,928	29,899	30,890	31,908	32,952	34,019
Padilla	10,444	10,645	10,847	11,049	11,252	11,456	11,660

Table 3.1: Estimated population annually aggregated per municipality from 2002 to 2008

To conclude the exploratory analysis using a set of potential non-climatic factors associated with dengue transmission, several tests were conducted via the Pearson coefficients. They included the variables of: Percentage of unoccupied households, Percentage of households that receive garbage collection service, Percentage of households with access to the piped water supply network, Altitude, Population density, Urban population, Percentage of households that receive electricity service and Percentage of households with domestic air conditioning equipment. To this end, Table 3.2 summarizes these outcomes:

Pearson Coefficient	
(1) % unoccupied	0.43
(2) % garbage collection	0.08
(3) % water supply network	-0.47
(4) Altitude	-0.25
(5) Pop density	-0.06
(6) Urban population	-0.05
(7) % electricity service	0.007
(8) % domestic air conditioning	0.07

Table 3.2: Pearson’s coefficients of potential non-climatic factors associated with the incidence rate of dengue fever: (1) Percentage of unoccupied households, (2) Percentage of households that receive garbage collection service, (3) Percentage of households with access to the pipe water supply network, (4) Altitude, (5) Population density, (6) Urban population, (7) Percentage of households that receive electricity service, (8) Percentage of households with domestic air conditioning equipment

3.5 Data exploration using climatic factors

As it was stated in chapter 2 (section 2.3.2) the initial data provided by the Ministry of Environment based in Zulia state, was limited due to missed values of continuous records reported by the local meteorological stations. However the urgent need for conducting epidemiological studies of dengue fever in Venezuela with cutting edge techniques lead us to integrate accurate satellite information for climatic factors in Zulia state.

To this end data exploration using temperature indexes and rainfall indexes in conjunction to the El Nino 3.4 (El Nino Southern Oscillation phenomenon) using remote sensing data (Kempler, 2013) was conducted in this section.

3.5.1 Data exploration using temperature indexes

As it was mentioned in previous chapter 2 (section 2.3.2) the temperature indexes were provided by the National Aeronautics and Space Administration (NASA) (Kem-

pler, 2013). They included: average, maximum and minimum surface temperature (expressed in $^{\circ}\text{C}$) from 2002 to 2008. They were monthly aggregated and allocated into four symmetric and non-overlapping grids. These grids were sized 2.0° latitude by 2.0° longitude covering the whole of Zulia state. According to the analysis, the average temperature in Zulia state fluctuates between 22.9°C to 29.7°C , providing an ideal condition for the development of *Aedes* species (Hii et al., 2009) and consequently the spread of the disease.

Figure 3.17 displays a time series of average surface temperature monthly aggregated over 2002 to 2008. The climatic data was plot across the four grids previously defined with latitude-longitude coordinates. This finding revealed a clear evidence that the warmest municipalities are located in the northern part of the state covering by GRID I and GRID II. The most affected municipalities being: Padilla, Miranda, Cabi-mas, Maracaibo, San Francisco, Lagunillas amongst others.

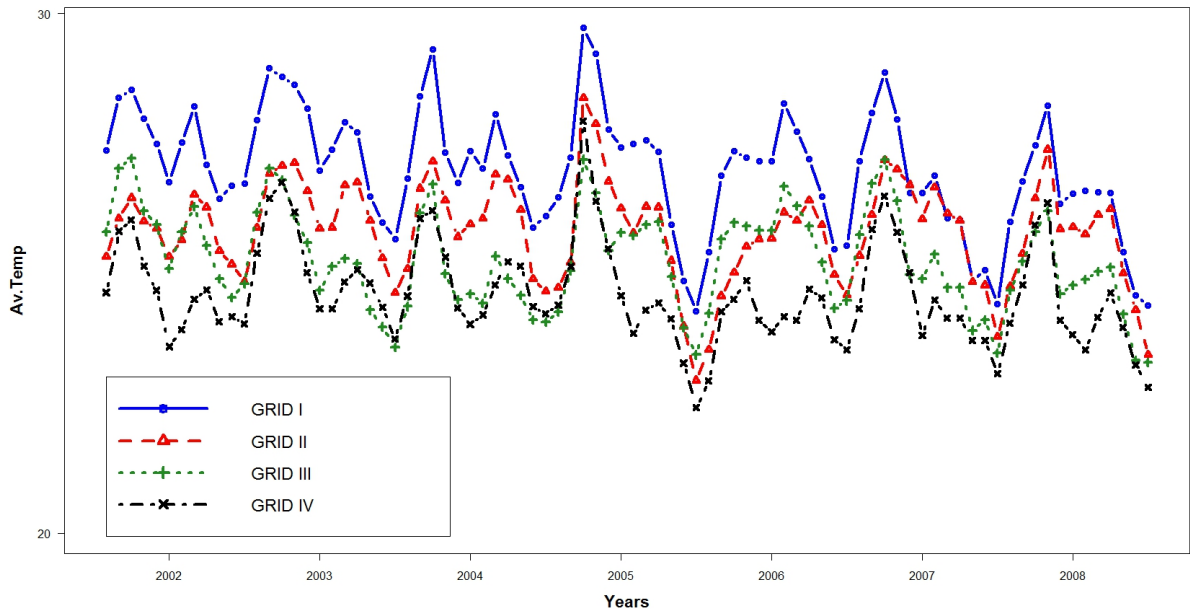


Figure 3.17: Time series of monthly aggregated average temperature per grids in Zulia state

It was also observed a seasonal pattern regarding hot conditions, during two different periods. The first hot season occurs during February to April and the second period from July to September. However these findings led us to explore the evolution of maximum and minimum temperatures, seeking greater variability across grids which could better justify a spatio-temporal analysis of dengue fever in Zulia state.

To this end Figure 3.18, shows the evolution of maximum temperature split into the four grids previously defined. As a result it was observed that maximum temperature per grid was widely spread, which might justify the inclusion of this index to capture the variation at the spatial dimension.

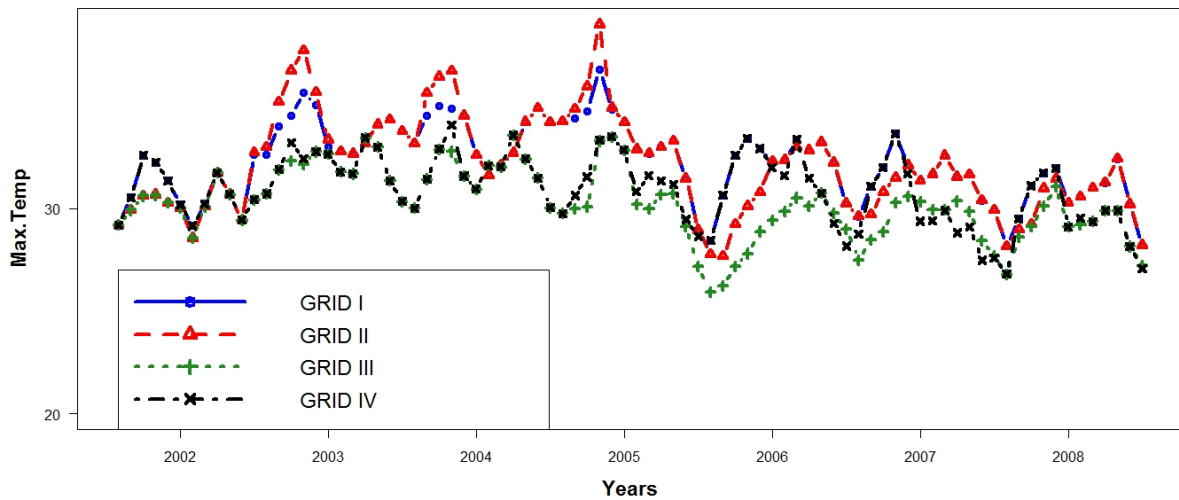


Figure 3.18: Time series of monthly aggregated maximum temperature per grids in Zulia state

This variation of maximum temperature settled by grids is summarized in Table 3.3. It is seen that GRID I and II experienced the peaks of maximum temperature, with levels between: 27.8°C and 38.9°C. The most affected locations being: Padilla, Miranda, Cabimas, San Francisco, Lagunillas amongst others.

Location	Temperature ranges
Grid IV	26.8°C to 34.0°C
Grid III	25.9°C to 33.5°C
Grid II	27.8°C to 36.7°C
Grid I	27.2°C to 38.9°C

Table 3.3: Maximum temperatures settled by GRID I, II, III, IV

Furthermore, Figure 3.19 displays a scatterplot of monthly aggregated maximum temperature averaged per grid versus count cases of dengue on the whole Zulia (covering 84 months) from 2002 to 2008.

Although several modelling studies regarding dengue transmission have focussed on the association with average temperature (Lowe et al., 2011; Chen et al., 2010) the preference in including maximum temperature for modelling purposes in Zulia state, Venezuela is because the severity of the disease reached a peak of 3,756 cases when the maximum temperature was around 30°C. But when the temperature was greater than 31°C the cases of dengue drastically diminished in Zulia. These findings were consistent with the associations found in previous studies conducted in different parts of the world, regarding the temperature indexes (Githeko et al., 2000; Hii et al., 2009; Johansson et al., 2009; Fuller et al., 2009). But even more biological studies in the field of dengue transmission claim a decrease in the incubation period when levels of temperature range from 35°C to 30°C (Brunkard et al., 2008). As a result mosquitoes become infected more quickly. These statements might justify the inclusion of maximum temperature instead of average or minimum temperature for modelling the evolution of dengue transmission in Zulia state.

Finally the variation of minimum temperature settled by grids were summarized in Table 3.4. It is seen that GRID I and II experienced the highest values of minimum temperature, with levels between: 18.24°C and 19.10°C located on the north of the state.

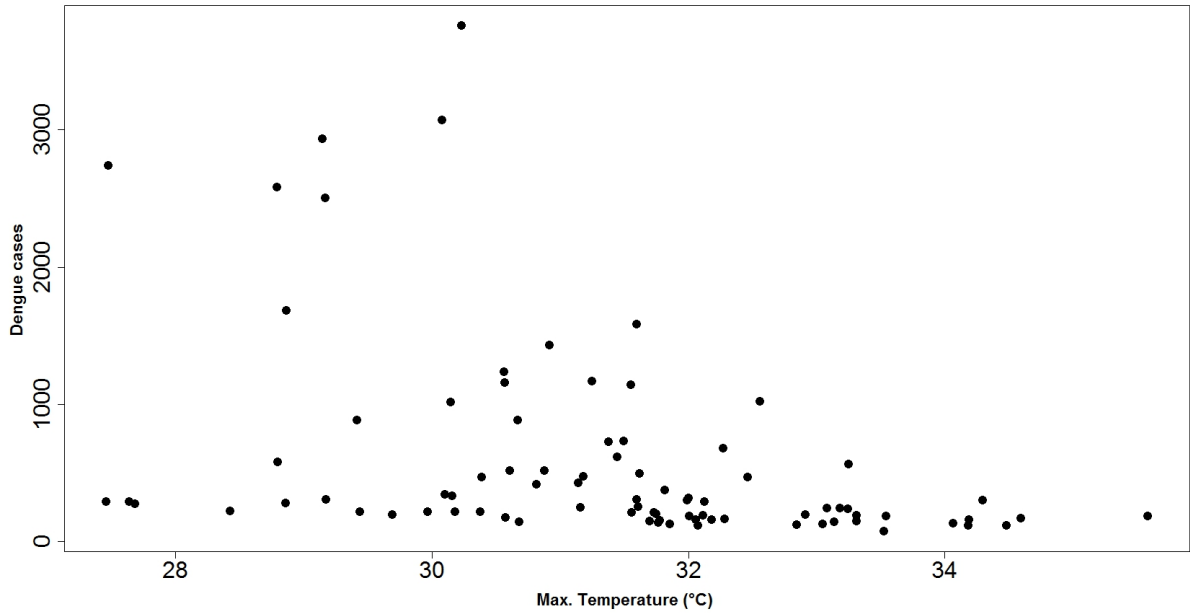


Figure 3.19: Scatterplot of monthly aggregated maximum temperature averaged per grid versus dengue cases in the whole of Zulia

Location	Min Temperature
Grid I	18.24°C
Grid II	19.10°C
Grid III	14.14°C
Grid IV	14.14°C

Table 3.4: Average of minimum temperature settled by GRID I, II, III, IV

However, elucidate which temperature index better captures the dynamics of dengue fever in Zulia state, shall be one of the tasks to be investigated in subsequent analyzes.

3.5.2 Data exploration using rainfall indexes

Moving on to the analysis of monthly rainfall rate, aggregated per grid in the whole of Zulia. A time series showing the evolution from 2002 to 2008 was displayed in Figure 3.20. These findings revealed a range of rainfall rate from 0 to 238.4 mm. The most affected municipalities with heavy rainfall being located on GRID III and GRID IV,

both in the south of Lake Maracaibo.

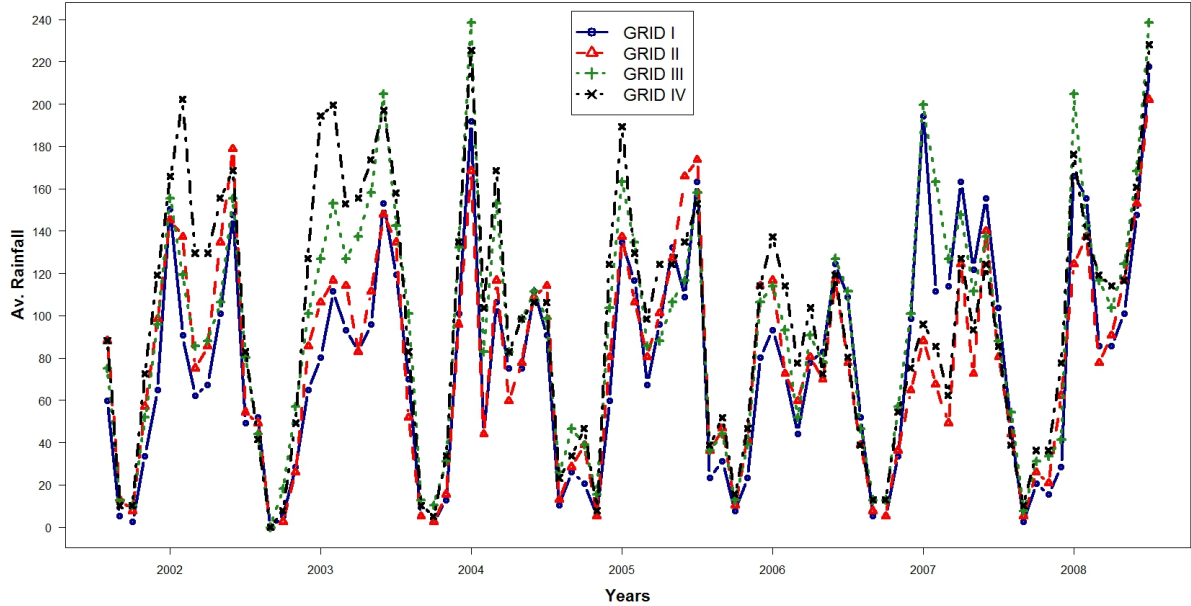


Figure 3.20: Time series of monthly aggregated average rainfall per grids in Zulia state

This figure also reveals a seasonal pattern with two major periods of intense rainfall ($> 100mm$). The first from May to June and the second from October to November. The highest peak of 238.4 mm was reached in 2004. This pattern confirms an inter-annual variation that needs to be fully understood in terms of dengue variation in Zulia state. On the other hand January, February and March were recorded as the driest months with low levels of rainfall ($< 15mm$).

The next plot was made to assess the potential relationship amongst average rainfall and cases of dengue fever in Zulia state. To this end Figure 3.21 displays a scatterplot of monthly aggregated average rainfall averaged per grid versus count cases of dengue on the whole Zulia (covering 84 months) from 2002 to 2008. Although again, it is difficult to reach a conclusion regarding the real shape of such association, based merely on this plot. It is seen that levels of rainfall greater than 80 mm cause drastically

diminution on dengue cases in Zulia state, whereas values between 0 to 80 mm, cause dengue cases highly noticeable. As stated above this complex association needs to be validated with more advanced techniques in subsequent analysis.

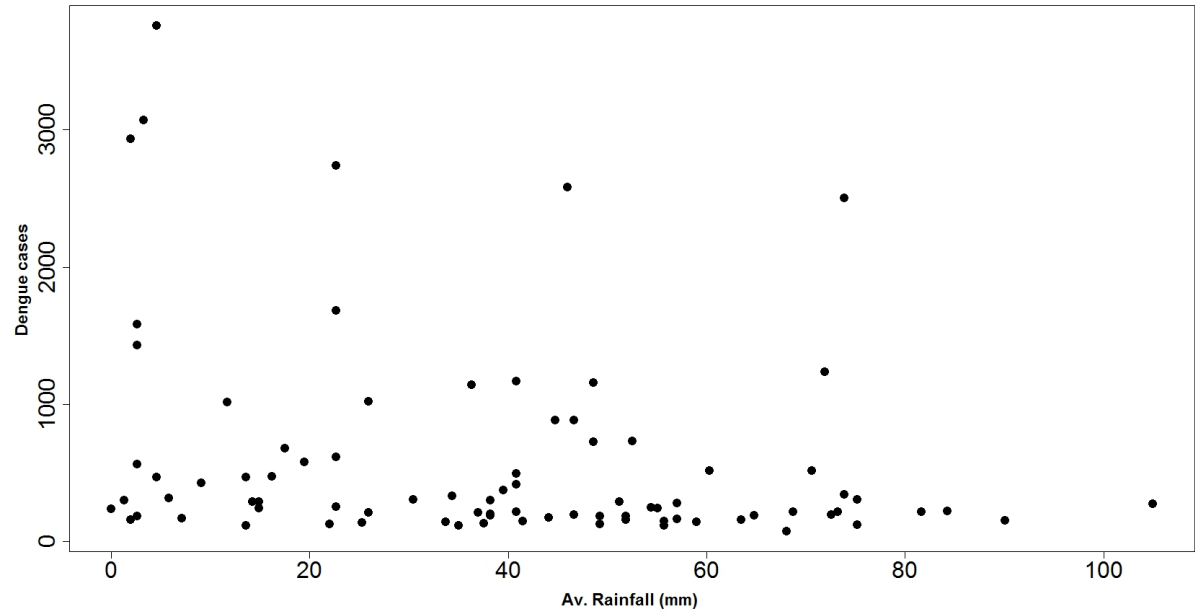


Figure 3.21: Scatterplot of monthly aggregated average rainfall averaged per grid versus dengue cases in the whole of Zulia

In a similar way, the maximum and minimum rainfall were plotted but no extra information was obtained. However assessing the appropriated rainfall index is a task that needs to be overcome. Finally it should be noted that even though some studies confirm a strong association between humidity and dengue transmission (Chen et al., 2010). These indexes were not included in the present analysis owing to homogeneity observed across the various grids that make up Zulia state.

3.5.3 Data exploration using the El Nino3.4 index

Towards an exploratory analysis of the El Nino Southern Oscillation (ENSO) associated with dengue cases within Zulia state, this researcher has proposed the use of six months delay and the El Nino3.4 as the best index that might capture the dynamic of the disease in the northwest part of Venezuela, from 2002 to 2008. Owing to a limited consensus regarding the appropriate time delay of the ENSO indices, associated with dengue transmission in different parts of the world (see Table 2.6). This decision was made based on a similar study associated with outbreaks of dengue fever, carried out in other Latin American countries (Lowe et al., 2011; Fuller et al., 2009). Some investigations undertaken in Brazil (Lowe et al., 2011) and Costa Rica (Fuller et al., 2009), revealed a negative effect on dengue transmission using six months delay of the El Nino3.4. Meaning that the colder extreme of the ENSO named La Nina heavily impacts the dynamic of the disease (Fuller et al., 2009).

Although only one initiative has been found which focus on the influence of the El Nino Southern Oscillation (ENSO) on a dataset from Venezuela (Rifakis et al., 2005). These findings showed no evidence of such association, perhaps because the time delay was not considered or because the statistical technique was not the most effective. As a result this limitation needs to be overcome in the present study, aimed to clarify the association of the ENSO and the dynamic of the disease in Zulia state, Venezuela.

Although this information was assumed to be constant across all the grids. Figure 3.22 displays a scatterplot of monthly aggregated the El Nino3.4 index versus count

cases of dengue fever on the whole Zulia (covering 84 months) from 2002 to 2008. It is seen a marked impact on the disease when values range from 27°C to 28°C , but a sharp decline when the El Nino 3.4 is lower than 27°C .

Finally, it is noteworthy that the observed trends based on the exploratory analysis, do not take into account the impact of several explanatory variables at the same time. As a result the implementation of more advanced techniques is the next step to be achieved.

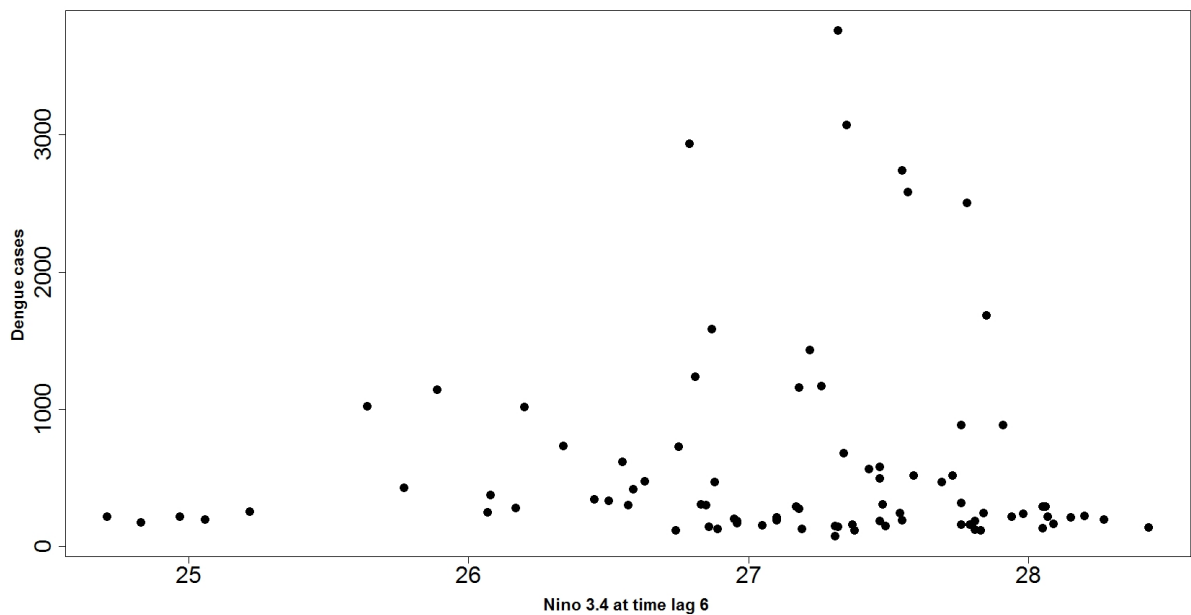


Figure 3.22: Scatterplot of monthly aggregated the El Nino3.4 (6 months lag delay) versus dengue cases in the whole of Zulia

3.6 Discussion and conclusions

This chapter comprises an extensive exploratory analysis underlying the most salient features that might explain the dynamic of dengue incidence in Zulia state, Venezuela fom 2002 to 2008. One of the greatest challenges to conduct epidemiological studies is

the limitation regarding the existence and quality of the dataset (Pfeiffer et al., 2008; Elliott and Wartenberg, 2004; Rothman et al., 2008). In this respect, Zulia state has not been an exception.

However the situation was overcome by gathering climatic data of temperature, rainfall and the El Nino Southern Oscillation indexes, via remotely sensed data provided by the National Aeronautics and Space Administration (NASA) (Kempler, 2013). On the other hand a set of data from different sources was aggregated on a regional basis, municipal level or the whole of Zulia and different temporal scales in accordance to the requirement of the exploration.

At first instance, the data was explored in a general fashion using the epidemiological dataset at the municipal jurisdictions from Zulia state. To this end, crude incidence rates of dengue fever were geographically visualized at municipal level, annually aggregated. As a result a wide spatial variation was observed during the years 2002, 2006 and 2007. Whilst the incidence rates during 2002, 2003 and 2005 seemed to be more homogeneous.

In addition critical levels of the severity of the disease were observed during 2007. Although the reasons for such pattern remain poorly understood until now. This chapter brings a set of evidences of potential explanatory variables associated with the transmission of the disease in Zulia state. To this end a wide range of categories were examined based on previous studies undertaken in different parts of the world (see section 2.5, chapter 2). They include: climatic, geographic, demographic, socioeconomic and individual factors.

An exploratory analysis for individual factors revealed that age group at risk was children between 5 and 14 years old. This finding confirms previous studies undertaken in Thailand and Florida (Kongsomboon et al., 2004; Morrison et al., 1998). The statement that middle childhood is the most vulnerable stratum of the population is a

fact that needs to be validated in subsequent analysis with more advanced statistical techniques.

In addition, the trends related to gender revealed a very similar pattern between both male and female groups. These findings were also consistent with some studies undertaken in others South American countries (Guha-Sapir and Schimmer, 2005).

This study also explored the potential linkage between a set of non-climatic factors and dengue transmission in Zulia state. To this end the average altitude expresses in metres above sea level and percentage of households with access to the pipe water supply network might be the variables that should be taken forward for examination in greater depth through more sophisticated modelling.

Pearson coefficients were also estimated amongst socioeconomic covariates to avoid conflict of collinearity (Johansson et al., 2009; Zuur et al., 2009; Zuur, 2012). To this end the percentage of households with access to the pipe water supply network seems to better capture the effect of the remaining covariates within the same category.

Regarding the climatic category, the exploratory analysis included: average, maximum and minimum temperature and average, maximum rainfall rate in conjunction to the El Nino3.4 index, monthly aggregated from 2002 to 2008. This study suggests that an estimation of optimal delay time between the climatic factors and the incidence rates of dengue fever in Zulia state (Rothman et al., 2008), needs to be elucidated in subsequent analysis. However a proposed delay time of: 0, 1 and 2 months for rainfall and temperature, and 6 months lagged effect for the El Nino3.4 index, are supported by previous evidence described in detail in chapter 2, section 2.5.2.

Moreover, seeking for evidence regarding the real shapes amongst climatic factors and dengue transmission, a set of scatterplots were performed. However although is difficult to reach a conclusion in this matter, based merely on visual plots. Some evidence

showed in most of the cases a complex pattern between epidemiological data of dengue and climatic factors. These patterns might be understood as non-linear associations.

These findings have been confirmed in similar studies undertaken in different parts of the world (Hii et al., 2009; Johansson et al., 2009; Fuller et al., 2009). These authors have stated that a non-linear association better capture the complex dynamic between dengue transmission and weather conditions.

Chapter 4

GLM: age and gender effects of dengue fever in Zulia state

4.1 Introduction

Previous epidemiological studies discussed in detail in chapter 2, revealed that children are the most vulnerable population in catching dengue fever when comparing findings from different parts of the world (Kongsomboon et al., 2004; Morrison et al., 1998; Raja and Devi, 2006; Dickin et al., 2013). Some researchers claim that it can be associated with low immunity in this stratum of the population (Kongsomboon et al., 2004), whereas others state that children are more exposed to open environments during play activities (Morrison et al., 1998), which increases the number of infected population.

The case study conducted in this thesis is based on epidemiological data of hospital admissions diagnosed with dengue fever aggregated by sex and age in the whole of Zulia from 2002 to 2008. A total of 24,861 cases grouped within the female category was weekly aggregated during 364 weeks from 2002 to 2008, and broken down into five groups of age: 0 to 4 years old, 5 to 14 years old, 15 to 24 years old, 25 to 44 years old and over 45 years old. In addition a total of 25,275 cases grouped within the male category was also split across the same groups of age.

To this end a Generalized Linear approach (GLM) for count data (Pfeiffer et al., 2008; Lovett and Flowerdew, 1989; Schwartz, 1993; Zuur et al., 2009) was designed to compare Relative Risk (RR) by exposure groups of age and sex using a dataset from the whole of Zulia. But although the main focus was originally on the development of a GLM using exposure factors of age and sex at municipality level, the limitations on epidemiological data at the named scales restricted the construction of such analysis. As a result, it was assumed that population size by age and gender for each municipality remained unchanged over the period of time under analysis.

The statistical framework developed in this chapter was based on a previous exploratory analysis discussed in detail in chapter 3. As it was explained, the exposure factors of age and gender in the whole of Zulia revealed that the group at risk was the children aged from 5 to 14 years old, followed by the stratum of the population aged from 15 to 24. As a result it was concluded that young population was the most vulnerable group in catching dengue fever in the whole of Zulia. Furthermore this exploratory analysis revealed that male and female's groups had a similar trend within both categories. In this respect the development of a GLM for count data in the whole of Zulia was driven by these findings.

Furthermore, the implementation of GLM for count data accounting for heterogeneity on the variance (Hoeff and Boveng, 2007; Zuur et al., 2009) was conducted via a frequentist approach and a Hierarchical Bayesian model by including a random term in the hierarchy, accounting for overdispersion (Zuur et al., 2009). The implementation of both perspectives, allowed the comparison of outcomes and methodologies (Ismaila et al., 2007) regarding the estimation on Relative Risk (RR) of dengue fever by exposure groups of sex and across the different categories of age: 0 to 4; 5 to 14; 15 to 24; 25 to 44 and over 45 years old in the whole of Zulia.

The computational implementation of the frequentist approach was based on the free statistical package R, version 2.15.2 (Zuur et al., 2009). And the computational

software for the Hierarchical Bayesian approach was the free package of WinBUGS, specialized on Bayesian inference (Gimenez et al., 2009; Abellan and Best, 2008; Lunn et al., 2000; Spiegelhalter et al., 2003).

4.2 Research objective

- Estimate Relative Risk (RR) of dengue fever in the whole of Zulia by comparing exposure groups of gender and age using a Hierarchical Bayesian approach and a frequentist model for comparative purposes.
- Compare limitations and benefits from both methodologies.

4.3 Research question

- Is the risk of dengue fever associated with a particular age group?
- Is the risk of dengue fever greater for male or female?

4.4 Model specification via a frequentist approach

At first instance a Quasi-Poisson approach was fitted due to the flexibility on the overdispersion parameter (Zuur et al., 2009) using count data of dengue cases in the whole of Zulia. The models at this stage included the population size used as an offset term in the whole of Zulia. These were monthly aggregated from 2002 to 2008 broken down in groups of age and gender. The groups of age, gender and years were specified within the models as categorical variables during the underlying period of time.

To this end a forward selection process via the lowest Akaike Information Criterion (AIC) was conducted to find the optimal fitted model (Zuur et al., 2009; Zuur, 2012). The set of models included main effects of the variables described above and some interaction terms amongst them, however the resulting structure included only main

effects of these covariates.

Moreover the dispersion parameter at this stage was estimated at $\hat{\phi} = 12.59$. According to some authors a large overdispersed parameter, justifies the use of a Negative Binomial (NB) model (Zeileis et al., 2008; Zuur et al., 2009) in a more formal fashion, with mean μ_i and shape parameter ϕ .

As a result a Negative Binomial approach for count dengue cases in Zulia state was performed in a frequentist way. The year 2002, group of age between 0 to 5 and female category were treated as baseline references. In addition, the extra Poisson variation might be based on spatial dependence amongst municipality units (Lawson, 2009; Banerjee et al., 2004) or because of a temporal dependence (Richardson et al., 2004) due to the lack of available data.

As a result the model specification via a frequentist approach is described below:

Model specification

$$count_i \sim NegBin(\mu_i, \phi)$$

$$\log(\mu_i) = \log(pop_i) + \beta_0 + \beta_1 predictor_1 + \beta_2 predictor_2 + \beta_3 predictor_3$$

$$\eta_i = \beta_o + \sum \beta_i predictor_i$$

where:

$\phi \rightarrow$ the dispersion parameter

$predictor_1 \rightarrow$ 2002, 2003, 2004, 2005, 2006, 2007, 2008

$predictor_2 \rightarrow$ female, male

$predictor_3 \rightarrow$ 0 to 4; 5 to 14; 15 to 24; 25 to 44; over 45

4.5 Analysis of results from the frequentist approach

The estimated Relative Risk (RR) of dengue fever in the whole of Zulia for every group of age and period of time monthly expressed, via a Negative Binomial (NB) regression using a frequentist perspective, revealed that children aged 5 to 14 years old were statistically significant with a 1.59 fold increased RR compared to the baseline group of 0 to 4 years old (see Table 4.1).

Negative Binomial			
Estimates	Freq. mean	Freq.RR	95% CI
Intercept	-5.92		
year 2003	-1.31	0.26	[0.23;0.30]
year 2004	-1.04	0.35	[0.31;0.40]
year 2005	-1.13	0.32	[0.28;0.36]
year 2006	0.001	1.00	[0.88;1.13]
year 2007	0.79	2.20	[1.95;2.48]
year 2008	-0.21	0.80	[0.71;0.91]
age group 5 to 14	0.46	1.59	[1.43;1.77]
age group 15 to 24	0.07	1.07	[0.96;1.19]
age group 25 to 44	-0.64	0.52	[0.46;0.58]
age group over 45	-1.18	0.30	[0.27;0.34]

Table 4.1: Negative Binomial estimates via a frequentist approach

Furthermore a large RR parameter estimated in 2007, showed that the risk of dengue fever increased by 2.20 times during this year, when comparing to the baseline category of year 2002. However the year 2006 did not emerge as being significantly different to that of the reference year. Moreover gender was dropped from the model because it proved to be non-significant.

In addition, these estimated parameters per group including the 95% Confidence Interval (CI's) can be visualized in Figure 4.1 by computing the standard error of the Negative Binomial frequentist approach.

The overall finding at this stage confirms the outputs obtained from the exploratory

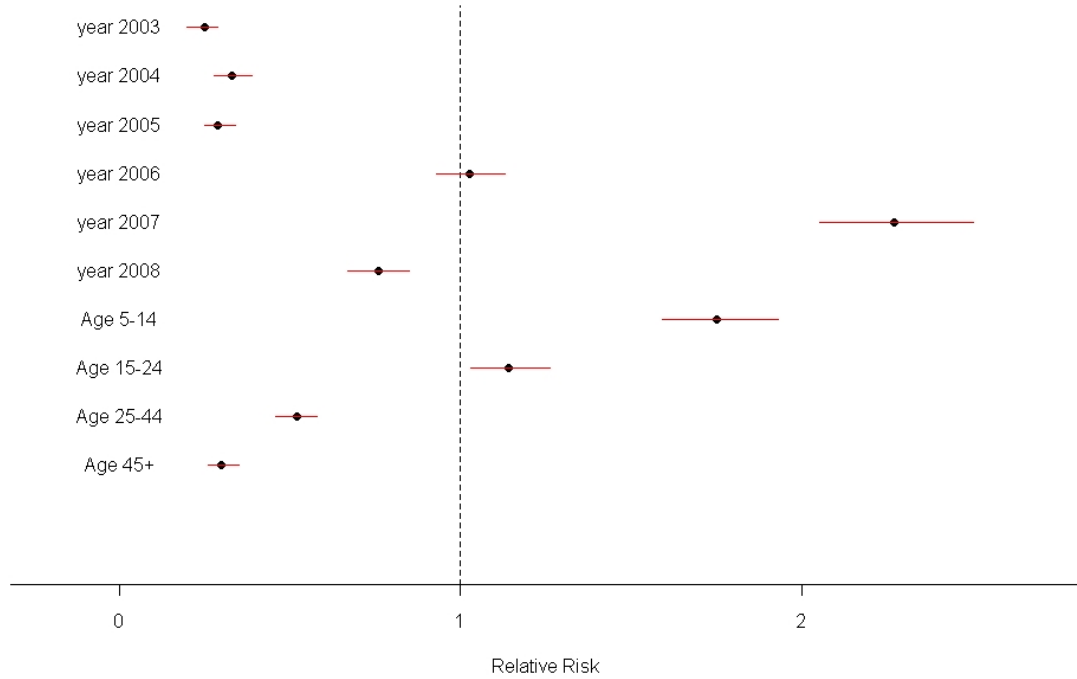


Figure 4.1: Confidence intervals (95%) of RR by age groups and years

analysis regarding groups of age, gender and years. Consequently the attention in the public health sector of Venezuela needs to be brought towards the young population from Zulia state. This stratum has been indicated as the most vulnerable group according to the present analysis. Thereby effective preventive programmes need to be targeted.

However a more accurate approach needs to be built accounting for space and time dimensions. At this respect a Spatio-Temporal modelling at municipality level might be the best option due to the availability of the data with higher resolution. Furthermore, the inclusion of new explanatory variables might substantially improve the fit of the model. These covariates could be based on previous findings from the exploratory analysis discussed in chapter 3, including perhaps: socioeconomic variables, geographical and/or climatic factors with its optimal lagged effects.

4.6 Model specification via a Hierarchical Bayesian approach

The subsequent analysis involves the application of a Hierarchical Bayesian approach based on count data of dengue cases in Zulia state. This approach provides the benefit of a full probability distribution for the parameters of interest by including the observed data as a likelihood function and prior beliefs for the unknown quantities. By using the hierarchical structure this approach can easily handle the overdispersion (Chien and Chen, 2009; Spiegelhalter, 1998; Greendland, 2006; McMahon et al., 2006; Hoeff and Boveng, 2007; Weiss, 2010).

The computational software at this stage was implemented via the statistical package WinBUGS (version in Window of Bayesian Analysis Using Gibbs Sampling) freely available (Spiegelhalter et al., 2003).

The analysis conducted in this section was based on non-informative or vague prior information in which the posterior distribution was expected to be dominated by the likelihood or observed data (Tan, 2001). To this aim the approach with the Bayesian perspective was adapted to capture the variability of the mean and variance, through the inclusion of the overdispersion parameter.

By assuming two-stages levels within the hierarchy the Bayesian approach was performed following two different structures:

First approach: Bayesian Hierarchical Poisson model

The first structure involves a simple normal distributed random effect into the linear structure of the Poisson distribution, named Bayesian Hierarchical Poisson model (Kery, 2010; Ntzoufras, 2009; Zuur et al., 2009). Experts claim that the most common

prior distribution representing flat or non-informative prior is assumed to follow a Normal distribution with mean zero and a very large variance (Weiss, 2010; Spiegelhalter, 1998). However because the parameterization of the Normal distribution in WinBUGS is defined in terms of precision, the desired flat distribution with large variance was implemented by assuming a very small precision.

Thereby the structure of a Bayesian Hierarchical Poisson model is specified as follows:

Likelihood:

$$Y_i \sim \text{Poisson}(\mu_i)$$

$$\log(\mu_i) = \log(\text{pop}_i) + \beta_0 + \beta_1 \text{predictor}_1 + \beta_2 \text{predictor}_2 + \text{eps}_i$$

where:

$\text{predictor}_1 \rightarrow 2002, 2003, 2004, 2005, 2006, 2007, 2008$

$\text{predictor}_2 \rightarrow 0 \text{ to } 4; 5 \text{ to } 14; 15 \text{ to } 24; 25 \text{ to } 44; \text{over}45$

Prior and hyper priors:

The random term eps_i was accounting for overdispersion. It was added within the Poisson structure, via a Normal distribution also called hyperparameter by getting one level up within the hierarchy (Zuur et al., 2009; Ntzoufras, 2009). And also leading to a Generalized linear Mixed model (GLMM) or Hierarchical model with non-informative prior (Zuur et al., 2009; Ntzoufras, 2009).

In a technical point of view the β 's parameters were defined as vague normal priors on the following way:

$$\beta \sim \text{Normal}(0, 0.000001)$$

And the variance component was normally distributed:

$$eps_i \sim Normal(0, eps.tau)$$

In which the second level on the hierarchy was defined as:

$$eps.tau \sim Gamma(0.001, 0.001)$$

For a better coverage the WinBUGS code can be found in the Appendix B. However, a second model was proposed using the Bayesian perspective by adapting a Negative Binomial approach (Weiss, 2008).

Second structure: Bayesian Hierarchical Negative Binomial model

An alternative model was also conducted exploring the Bayesian framework by adapting the Negative Binomial formulation via a Poisson-Gamma distribution structure (Weiss, 2008; Ntzoufras, 2009). In this case the Poisson distribution was assumed to be a product of two parameters. Hence the parameterization turns out (Y_i) the reported cases of dengue fever during month i , with mean μ_i and dispersion parameter τ_i . As a result the specification of the model was defined as follow:

Likelihood:

$$Y_i \sim Poisson(\tau_i \mu_i)$$

$$\log(\mu_i) = \log(pop_i) + \beta_0 + \beta_1 predictor_1 + \beta_2 predictor_2 + eps_i$$

where:

$predictor_1 \rightarrow 2002, 2003, 2004, 2005, 2006, 2007, 2008$

$predictor_2 \rightarrow 0 \text{ to } 4; 5 \text{ to } 14; 15 \text{ to } 24; 25 \text{ to } 44; \text{over}45$

Prior and hyper priors:

The vague normal priors for the β 's parameters were defined as:

$$\beta \sim Normal(0, 0.000001)$$

And the dispersion parameter defined as:

$$\tau_i \sim Gamma(\alpha, \alpha)$$

$$alpha = exp(logalpha)$$

$$logalpha \sim Normal(0, 0.000001)$$

The code of WinBUGS can also be found in Appendix B.

4.7 Analysis of results from the Bayesian approach

After running the Hierarchical Bayesian Poisson regression via WinBUGS with 100,000 iterations as a burn-in period, the point estimations were summarized in Table 4.2. These estimates based on the posterior means were listed from b[1] to b[10]. The parameters coded between b[1] to b[6] represent the years: 2003, 2004, 2005, 2006, 2007 and 2008. In which year 2002 was considered as the baseline group. And from b[7] to b[10] the groups of age were covered as: 5 to 14, 15 to 24, 25 to 44 and over 45 years old respectively with 0 to 4 being the baseline group.

As a result, the findings revealed that children aged from 5 to 14 years old had 1.57 fold increased Relative Risk (RR) when compared to the baseline category aged from 0 to 4 years old. In a similar way the year 2007 showed that Relative Risk increased by 2.21 times when compared to the baseline year 2002. These results were consistent with the findings from the frequentist approach detailed in section 4.5.

However to validate these outcomes some plots were provided to visualize the posterior distributions of the parameters under analysis. First of all, the posterior densities for the parameters: year 2003 defined as $\text{beta}[1]$, year 2004 defined as $\text{beta}[2]$ and group of age from 5 to 14 years old denoted as $\text{beta}[7]$ showed clearly smooth and unimodal shapes, with posterior point means of: -1.3, -1.03 and 0.45 respectively (see Figure 4.2).

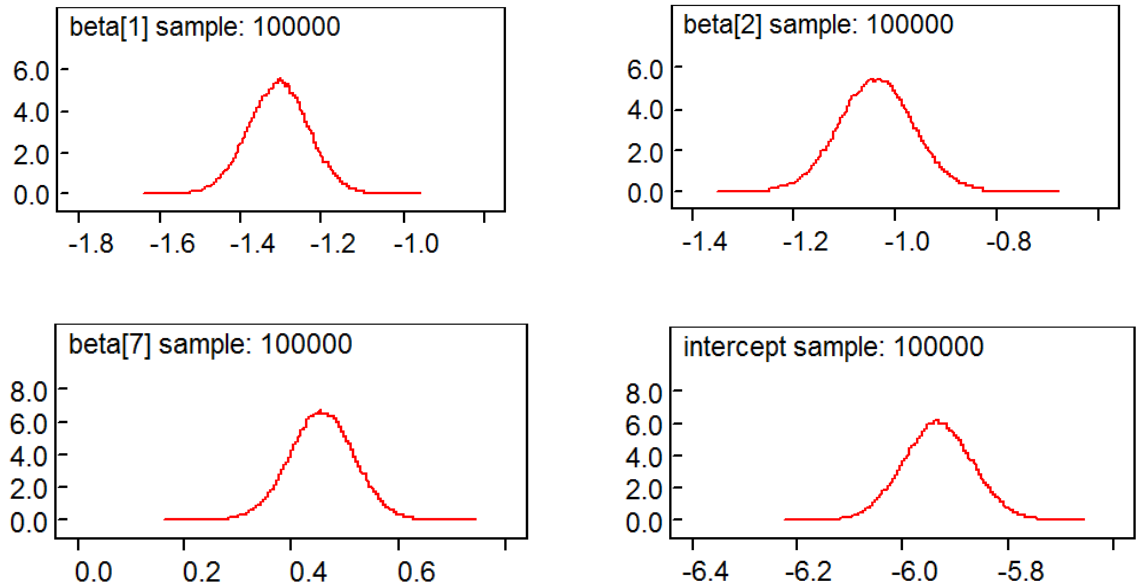


Figure 4.2: WinBUGS outputs showing the Posterior kernel densities of $b[1]$ year 2003, $b[2]$ year 2004 and $b[7]$ group of age from 5 to 14 years old.

It was showed that the remaining parameters had a similar pattern. Furthermore, the trace patterns of the estimates parameters against the number of iterations were

plotted in Figure 4.3. It was seen quite dense chains, meaning a good mixing of the parameters. In addition, these chains appear to have reached a stationary distribution in which the centered point is the mean of the posterior distributions.

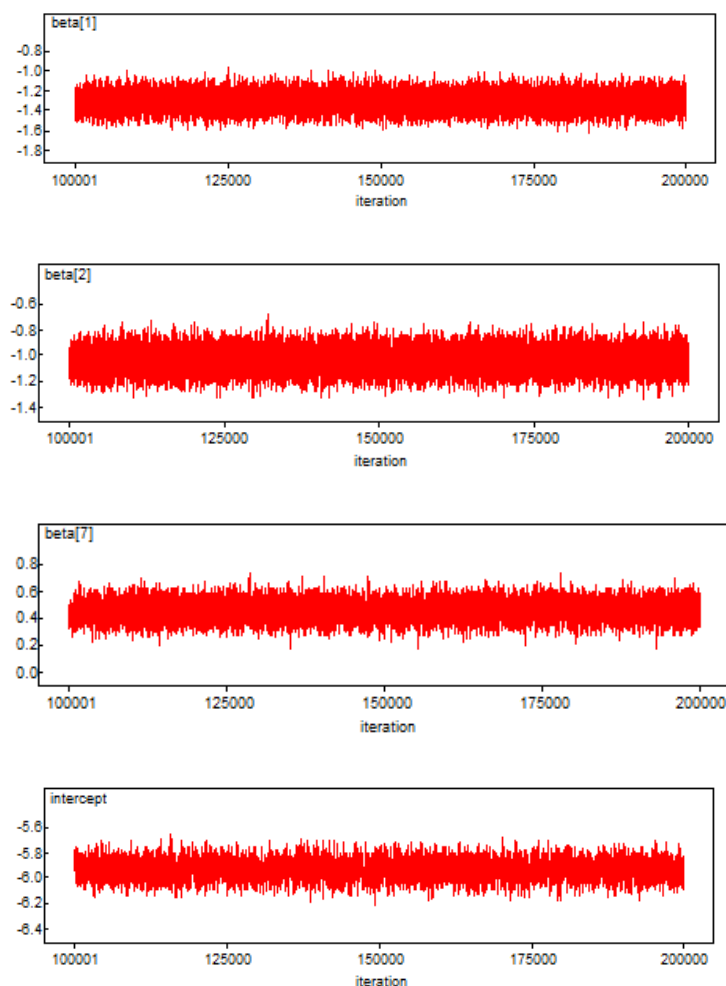


Figure 4.3: WinBUGS outputs showing the Dynamic traces of $b[1]$ year 2003, $b[2]$ year 2004 and $b[7]$ group pf age from 5 to 14 years old.

It was observed as well that the remaining parameters had a similar pattern. Moreover, the autocorrelation functions (see Appendix B) of every parameter indicates that the posterior distribution mixed slowly, which is a good pattern in term of correlated values. And finally the visual diagnostic revealed that the bivariate posterior scatter

plots around the mean was randomly distributed for all possible combination of parameters (see Appendix B). These patterns revealed that the convergence was reached.

As was mentioned in section 4.6 a second structure was also specified using a Bayesian Hierarchical Negative Binomial approach. It was seen that this structure was more consistent with the frequentist model discussed on section 4.5. To this end the new point estimators generated by the posterior mean of the Negative Binomial Hierarchical Bayesian version were displayed in Table 4.3, in which the standard error was based on its posterior distribution. Consequently the Credible Interval (CrI) at 95% was estimated.

Node	Bayesian Mean	Bayesian RR	95% CrI
Intercept	-5.93		
beta[1]=year 2003	-1.30	0.27	[0.23;0.31]
beta[2]=year 2004	-1.03	0.35	[0.30;0.40]
beta[3]=year 2005	-1.13	0.32	[0.27;0.37]
beta[4]=year 2006	0.002	1.00	[0.87;1.15]
beta[5]=year 2007	0.79	2.21	[1.92;2.54]
beta[6]=year 2008	-0.21	0.81	[0.70;0.93]
beta[7]=age group 5 to 14	0.45	1.57	[1.39;1.77]
beta[8]=age group 15 to 24	0.067	1.07	[0.94;1.20]
beta[9]=age group 25 to 44	-0.64	0.52	[0.46;0.59]
beta[10]=age group 45 or more	-1.19	0.30	[0.26;0.34]

Table 4.2: Posterior summary of the Bayesian Hierarchical Poisson approach using WinGUBS

NB Frequentist-using R				
Estimates	mean	RR	95% CI	
Intercept	-5.92	—	—	
year 2003	-1.31	0.26	[0.23;0.30]	
year 2004	-1.04	0.35	[0.31;0.40]	
year 2005	-1.13	0.32	[0.28;0.36]	
year 2006	0.001	1.00	[0.88;1.13]	
year 2007	0.79	2.20	[1.95;2.48]	
year 2008	-0.21	0.80	[0.71;0.91]	
age group 5 to 14	0.46	1.59	[1.43;1.77]	
age group 15 to 24	0.072	1.07	[0.96;1.19]	
age group 25 to 44	-0.64	0.52	[0.46;0.58]	
age group 45 or more	-1.18	0.30	[0.27;0.34]	

NB Bayesian-using WinBUGS				
Estimates	mean	RR	95% CrI	
Intercept	-5.92	—	—	
beta[1]=year 2003	-1.31	0.26	[0.23;0.31]	
beta[2]=year 2004	-1.04	0.35	[0.30;0.40]	
beta[3]=year 2005	-1.13	0.32	[0.27;0.37]	
beta[4]=year 2006	0.001	1.00	[0.87;1.15]	
beta[5]=year 2007	0.79	2.20	[1.99;2.53]	
beta[6]=year 2008	-0.21	0.80	[0.70;0.92]	
beta[7]=age group 5 to 14	0.46	1.59	[1.41;1.79]	
beta[8]=age group 15 to 24	0.072	1.07	[0.95;1.21]	
beta[9]=age group 25 to 44	-0.64	0.52	[0.46;0.59]	
beta[10]=age group 45 or more	-1.18	0.30	[0.27;0.34]	

Table 4.3: Comparison between both the Frequentist NB approach and the Bayesian NB approach

Comparing similarity of outcomes between both NB approaches was a good practice within this study. As it was mentioned by experts, exploring different options for the prior distributions of the unknown quantities could lead to further studies in this field. As a result some authors advise the use of Uniform prior distribution or the inverse Gamma as potential alternatives (Gelman, 2006).

4.8 Comparing limitations and benefits from both methodologies

The main interest in this chapter was modelling the exposure risk factors of age, gender and years using a dataset from the whole of Zulia state, Venezuela from 2002 to 2008. However the findings presented in this chapter were based on a comparison between the Bayesian perspective and the frequentist approach accounting for heterogeneity on the variance.

In this context, most of the authors in the field of applications prefer the use of the Bayesian perspective (Trevisani and Torelli, 2007; Gimenez et al., 2009; Basanez et al., 2005; Tan, 2001; Chien and Chen, 2009; McMahon et al., 2006; Lowe et al., 2011) for the following reasons:

1. The specification of the model can be straightforwardly defined within the Bayesian framework.
2. The growing development of computational software via the implementation of Markov Chain Monte Carlo (MCMC) algorithms.

Although all of these statements seem to favour the Bayesian perspective, the construction of a Spatio-Temporal modelling via the frequentist approach was the analysis subsequently conducted in this project. The reasons that support such decision will be

discussed below.

The subsequent analysis is an extension of the Generalized Linear Mixed model (GLMM) called Generalized Additive model (GAM) (Wood, 2006). The Spatio-Temporal structure proposed involves the implementation of a semiparametric approach for the incidence rates of dengue fever in Zulia state. This methodology will include smooth terms for the climatic factors and the inclusion of a set of non-climatic covariates, monthly aggregated at municipality level.

Because of a lack of technical reports in the field of epidemiological studies. A more systematic approach is suggested using the epidemiological data of dengue cases from Zulia state. To this end a frequentist approach using the GAM structure was preferred due to complex model selection and validation processes (Wood, 2006; Zuur, 2012) that needed to be assessed using a simpler structure.

4.9 Discussion and conclusions

The study conducted in this chapter consists of the estimation of Relative Risk (RR) of dengue fever in the whole of Zulia by comparing exposure groups of gender, age and years. This study was based on epidemiological data of hospital admissions diagnosed with dengue fever aggregated by sex and age in the whole of Zulia. To this end a total of 24,861 female cases and 25,275 male cases were grouped into five age groups: 0 to 4 years, 5 to 14 years, 15 to 24 years, 25 to 44 and over 45 years. However a limitation owing to a lack on epidemiological data aggregated at the mentioned exposure groups of age and gender, restricted the construction of a similar approach using exposure factors at municipality level. Under such circumstances, it was assumed that age structure per gender remained unchangeable at municipality level over the period under analysis. This assumption might be a cause of a potential overdispersion issue.

The study was initially motivated by the findings from the exploratory analysis discussed in detail in chapter 3. To answer the research questions proposed in this chapter, the structure of the models was based on a Generalized Linear approach (GLM) for count data (Pfeiffer et al., 2008; Lovett and Flowerdew, 1989; Schwartz, 1993; Zuur et al., 2009) aimed to compare Relative Risks (RR) by exposure groups of age, sex and years in the whole of Zulia. The implementation of the GLM's for count data accounting for heterogeneity on the variance (Hoeff and Boveng, 2007; Zuur et al., 2009) was conducted via a frequentist approach and a Hierarchical Bayesian model by including a random term into the hierarchy (Zuur et al., 2009). Thereby a comparison of outcomes and methodologies (Ismaila et al., 2007) was performed.

In the first instance, a set of Quasi-Poisson models were fitted because of the flexibility on the overdispersion parameter (Zuur et al., 2009). At this stage, a forward selection process via the lowest Akaike Information Criterion (AIC) was conducted (Zuur et al., 2009; Zuur, 2012) and a large overdispersion parameter was found in 12.59. This value justified the use of a more accurate approach of Negative Binomial for subsequent analysis. However the extra Poisson variation might be caused by a spatial dependence amongst municipality units (Lawson, 2009; Banerjee et al., 2004) but this statement could not be justified at this stage due to a lack of available data. Consequently the analysis involved the application of a Negative Binomial model based on count data of dengue cases via a frequentist and a Hierarchical Bayesian perspective. The implementation of both versions allowed the comparison of outcomes and methodologies (Ismaila et al., 2007) regarding the estimation on Relative Risk (RR) of dengue fever across the different categories of age: 0 to 4, 5 to 14, 15 to 24, 25 to 44 and over 45 years old, over the whole of Zulia.

The comparison between both NB approaches showed that children aged from 5 to 14 years old had a 1.59 fold increased Relative Risk (RR) when compared to the baseline category aged from 0 to 4 years. The severity of the disease was also observed within the stratum of the population from 15 to 24 years old. These outcomes

confirmed that group at risk in Zulia state, Venezuela were children and youth population. This statement was validated by similar studies conducted in different parts of the world in which the highest reported cases were observed amongst the youngest (Kongsomboon et al., 2004; Morrison et al., 1998; Raja and Devi, 2006). Given the opportunity of implementing strategies for the control and prevention of dengue fever in Zulia state, an effective initiative is needed involving scientists, health agencies and the local community aimed to at coordinating an integrated vector programme. In this context, the WHO made a proposal for disseminating the information about dengue transmission in order to educate people about the techniques that could help to protect or prevent the infection (Chan, 2012). The WHO also recommends effective intervention to reorganize the public health centres to provide the people with timely access to them during outbreaks of dengue fever (Chan, 2012).

The findings of this study also showed that the year 2007 had the largest Relative Risk (RR) increased by 2.20 times when compared to the baseline year 2002. Detailed examination of the outcomes also showed that the previous year (2006) had a high Relative Risk factor, but it was low in the preceding years, 2003, 2004 and 2005. This pattern leads us to investigate the reasons for the evolution of dengue fever transmission over time. It has been argued that epidemics of dengue fever are followed by endemic cycles periodically observed every 3 to 5 years (Bennett et al., 2009). The evolution of dengue fever in Zulia state from 2002 to 2008 confirmed such an assertion.

In addition, although these findings were based on a NB Hierarchical Bayesian model using non-informative or vague prior information in which the posterior distribution was dominated by the likelihood or observed data (Tan, 2001). More studies are needed to explore potential alternatives of prior distributions in future studies (Gelman, 2006). On the other hand, experts in the field of applications emphasize the advantages of the Bayesian framework, because of the straightforward specification of the models and because of the growing development of computational software via the Markov Chain Monte Carlo (MCMC) algorithms, which facilitates the implementation

of applied studies (Trevisani and Torelli, 2007; Gimenez et al., 2009; Basanez et al., 2005; Tan, 2001; Chien and Chen, 2009; McMahon et al., 2006; Lowe et al., 2011). Finally uncertainties over epidemiological data on the type of serotypes circulating during the period of time under analysis, was a limitation within the surveillance information provided in Zulia state. Experts postulate a temporal cycle of serotypes circulating in geographical contexts (Bennett et al., 2009) which could have a direct effect across the various strata of the population categorized by age. Moreover a deeper analysis is needed to explore some other factors associated with the spread of the disease in Zulia, looking for spatial and temporal dimensions in dengue transmission. This analysis was widely covered in chapter 5.

Chapter 5

GAM: Spatio-temporal modelling of dengue fever, towards an EWS

5.1 Introduction

The exploratory analysis discussed in detail in chapter 3, revealed a complex linkage between climatic factors and the occurrence of dengue fever in the Northwest part of Venezuela that will be elucidate in this chapter. Although a similar epidemiological study conducted in Brazil (Lowe et al., 2011) agreed with this statement, a linear parametric assumption was performed in the referred study via a Generalized Linear Mixed model (GLMM). As a result some limitations were found in modelling the real dynamic of dengue transmission.

Some other investigations performed in different parts of the world have confirmed a non-linear pattern between weather conditions and dengue fever (Hii et al., 2009; Johansson et al., 2009; Fuller et al., 2009). However for the first time, an application implementing a semi-parametric approach of the incidence of dengue fever in Zulia state has been proposed in this study by extending the GLMM structure into a novel approach called Generalized Additive model (GAM) (Hastie and Tibshirani, 1986). However the reduced number of technical reports in this field became the greatest limitation to be overcome.

From a technical point of view, the final implemented GAMM structure allowed the inclusion of unknown smooth functions (Hastie and Tibshirani, 1986) that captured the complexity of the associations amongst the covariates and the response variables tailored to explain the real dynamic on dengue transmission in Zulia state. This structure was based on smooth terms for climatic data and a set of non-climatic covariates in a categorical form. All the predictors were monthly aggregated across the 21 municipalities that comprise Zulia state from January 2002 to June 2009, accounting for both spatio and temporal dimensions. In addition the municipality levels were treated as a random effect by extending into a GAMM framework (Wood, 2006).

Seeking the impact that climatic factors have on the incidence of the disease, the data originally provided by the local meteorological stations of Zulia state was replaced by a set of satellite remote sensing climatic information provided by the National Aeronautics and Space Administrations (NASA) (Kempler, 2013). This information was ideal for the construction of a Spatio-Temporal model with a high standard of accuracy.

Another important issue handled in this chapter was the estimation of optimal lagged effects (Rothman et al., 2008) for the climatic factors of rainfall, temperature and the El Nino3.4. Although some lagged effects discussed in detail in chapter 2, were found in different geographical contexts. An empirical estimation of optimal lagged effects in Zulia state was carried out via a model selection of GAMs accounting for temperature and rainfall indexes at 0, 1 and 2 months delay and 6 months delay for the global climatic index of the El Nino3.4. This later delay time was previously validated in Brazil (Lowe et al., 2011).

Predicting dengue cases in Zulia state was also an analysis which fully merited attention in this chapter. Although some studies have been aimed at early detection of malaria (Bouma et al., 1997; Bouma and Van der Kaay, 1996; Abeku et al., 2004; Thomson et al., 2004) only few of them have focussed on outbreaks of dengue fever

(Lowe et al., 2013; Hii et al., 2012; Descloux et al., 2012; Hu et al., 2010). The importance of early detection for infectious diseases has been acknowledged by the WHO (Chan, 2012) as a powerful tool to lead governmental authorities for effectively managing interventions for the control and prevention of the diseases. In this context, a predictive GAMM was implemented for the first time after a comprehensive model selection process. Validation of such predictions was also assessed for a good performance using out-of-model data (Rodier et al., 2004; Hii et al., 2012; Lowe et al., 2013) from July 2009 to June 2010. As a result, the findings revealed a 90% confidence interval, which fulfilled the protocols requirements for the construction of an EWS with a high level of accuracy (Rodier et al., 2004; Connolly, 2005).

5.2 Research objectives

- Investigate the optimal lagged effects and climatic covariates associated with incidence rate of dengue fever in Zulia state.
- Develop a Spatio-Temporal model via a Generalized Additive framework using municipalities as a random effect.
- Develop a predictive model of dengue cases in Zulia state, with high standard of accuracy to be extended into an Early warning system.

5.3 Research questions

- Can the model robustly reveal the underlying relationship between the explanatory covariates and the response variable to confirm the biological principles involved with dengue transmission?
- Can a predictive model of dengue cases be developed with a high level of accuracy using an epidemiological dataset from Zulia state, Venezuela? Could this model be extended in future studies into an Early warning system including forecasts of dengue fever?

5.4 Model specification

A systematic approach was implemented in this section to illustrate the evolution in building a spatio-temporal GAMM structure for predictive purposes of dengue fever in Zulia state. This approach was conducted to investigate potential patterns that could contribute with the final version for predictive purposes. Although the process of construction relied on three different stages, each with its own model specification. These stages complemented each other towards the final version.

The computational analysis was based on the statistical software R and the package *mgcv* designed by S.Wood (Wood, 2006). The selection of this package was due to the versatility of the tool (Wood, 2006; Zuur et al., 2009). And the model specification at each stage are described below:

First stage, was conducted to investigate the optimal lagged effect and the climatic covariates associated with the incidence rates of dengue fever in Zulia state. A GAM structure was implemented using region as a factor. Space and time were considered in this analysis using monthly aggregated cases over the period from January 2002 to December 2008, across the five regions that comprise Zulia including: Guajira, East Coast, Capital, Perija and South. The inclusion of region at this stage, arose because previous evidence on the variation of health response across region was found in the exploratory analysis discussed in chapter 3. This approach was conducted following a Generalized Additive model (Hastie and Tibshirani, 1986) through a frequentist estimation procedure (Wood, 2006).

A set of models were built accounting for all possible combinations of the climatic covariates listed in Table 5.1 with different lagged effects. All these options included smoothed terms for: maximum and average temperature in conjunction with average, maximum and minimum rainfall with lagged effects from 0,1 and 2 months and the El Nino3.4 included at 6 months delay.

Covariate	lagged effect
max temp (°C)	0, 1 and 2 months
av temp (°C)	0, 1 and 2 months
max rainfall (mm)	0, 1 and 2 months
min rainfall (mm)	0, 1 and 2 months
av rainfall (mm)	0, 1 and 2 months
Nino 3.4	6 months

Table 5.1: Climatic covariates with lagged effects to be modelled

In addition, smoothed terms for a set of non-climatic covariates including: average altitude, percentage of households with access to the piped water supply network and area, expressed in square kilometers were aggregated per regions. An offset term for the population size was also included within the models in order to adjust for differences in the population size across the underlying regions (Terzi and Cengiz, 2009).

At first instance the GAM models were built via a Quasi-Poisson version. However after estimating an overdispersion parameter valued in 25.3 (Cleophas et al., 2010). The finding led to the construction of Negative Binomial version as the best option due to a large heterogeneity on the variance (Zuur et al., 2009). Through the Negative Binomial models, free restriction on the degree of freedom for the smooth terms was compared with some fixed values defined at $k = 3$ and $k = 5$ respectively. Fixed parameters of the degree of freedom were suggested by some authors as a good practice when dealing with a large number of observations (Zuur et al., 2009).

A forward selection process was conducted via the lowest Akaike Information Criterion (AIC) using the Negative Binomial GAMs models for count data (Wood, 2006) of dengue cases in Zulia state. However before running the GAMs structures, the phenomenon of multicollinearity was carefully checked (Berg, 2005; Guisan et al., 2002) in this study regarding the non-climatic covariates. As it will be discussed later, the finding at this stage revealed a clear evidence that maximum temperature and average rainfall correctly captured the dynamic of dengue transmission in Zulia state during 2002 to 2008. The El Nino3.4 at 6 months lagged effects in conjunction with lagged

effect at 1 month for the remaining climatic factors were found as the optimal options with a free restriction on the degree of freedom (see Analysis of results, section 5.5).

Second stage, some matters handling models in the first stage, led to the construction of a more accurate spatio-temporal modelling of dengue fever in Zulia state. Some of the reasons for improving the previous model was due to a visual evidence of distorted shapes regarding the smoothed estimated functions of climatic factors associated with the incidence rates of dengue fever across the different regions of: Guajira, East Coast, Capital, Perija and South (see Figure 5.1 in section 5.5). Another remarkable reason is the fact that the non-climatic factors of: altitude, area and percentage of households with access to the water piped supply network, treated as smoothed terms, might not be the best option due to a gap observed amongst the X-axis (see Figure 5.1 in section 5.5). Consequently these patterns might be understood as inaccuracy on the function's shapes. Thereby the non-climatic covariates might be implemented in a categorical form in this second stage. In addition the inclusion of municipalities instead of regions could be a better practice based on evidences of homogeneity from the residual plots (see Appendix C).

As a result a new spatio-temporal Generalized Additive Mixed model (GAMM) based on random effect for geographical entities of municipalities in Zulia state was proposed. These geographical entities were chosen because it was the finest georeferenced data available. According to experts a random effect into the previous GAM model was suggested due to the low number of municipalities (Wood, 2013), which allow for unobserved heterogeneity at this scale. And seeking for spatial autocorrelation the use of a finer resolution at municipal level was desired.

Furthermore, the models at this stage are characterized by the inclusion of additional terms which might allow the exploration of new outcomes that could benefit the modelling for predictive purposes. To this end, a new smoothed term was included within this structure allowing for a long term trend (Staelens et al., 2012; Under-

wood, 2009) through a variable named *timeserie*. This variable was mainly included to capture the long term trend by allowing the unknown covariates to drive the process (Staelens et al., 2012; Underwood, 2009). Thereby the data was monthly aggregated at municipal level, being *timeserie* enumerated from 1 to 84 covering from January 2002 to December 2008. In addition a second smoothed term for seasonal trend was also included within the model, represented by a cyclic argument using the *mgcv* package (Wood, 2006).

Moreover the set of non-climatic covariates were implemented in a categorical form. For a practical point of view the covariates percentage of households with access to the piped water supply network and altitude were grouped into three different categories. The levels of percentage of households with access to the piped water supply network were coded as follows: Coded 0, covering from 0 to 50% of the households with access to the water supply network. Coded 1, covering from 51 to 80% of the households. And Coded 2, from 81 to 100% of the households.

On the other hand the categorical groups of altitude were coded as: Coded 0, alluvial plain landscape which are composed by heights from 0 to 600 metres above sea level. Coded 1, middle elevations with heights from 601 to 1,500 metres. And Coded 2, highlands from 1,501 to 3,000 metres. These levels were defined following the geographical features of Zulia state. Coded 0 for instance, captures the coastal zone along the Gulf of Venezuela, marked mainly by alluvial plain landscape. This zone is also characterized by being semi arid with dry and warm conditions. Coded 1, covers the middle elevations in Zulia state. And Coded 2, covers the highlands mainly located on the Serrania de Perija.

An argument gamma defined equal to 1.4 ($\gamma \approx 1.4$) was also included to avoid overfitting issues (Wood, 2006). And finally a fixed term of $k=5$ was implemented for a better performance of the smoothed terms. As a result, the model specification at the second stage was defined as follows:

$$count_{it} \sim NegBinomial(\rho, \mu_{it})$$

$$\begin{aligned} \log(\mu_{it}) = & \log(pop_{it}) + f_1(x_{it}) + f_2(z_{it}) + f_3(w_t) \\ & + f_4(month_t) + f_5(timeserie_t) + Zb_i + X_i^* \beta^* \end{aligned}$$

where:

$count_{it}$ → dengue counts with $i=1,...,21$ and $t=1,...,84$. Being January month 1

μ_{it} → mean of dengue count

ρ → scalar parameter of the NB

$\log(pop_{it})$ → population size per municipality

x_{it} → mean rainfall at 1 month delay

z_{it} → maximum temperature at 1 month delay

w_t → Nino3.4 index at 6 months delay

Z → random effect. Matrix nested within municipality

b_i → random effect vector

$X_i^* \beta^*$ → parametric part, comprised by altitude and percentage of households with piped water supply as categorical levels

Third stage, although the findings from the second stage revealed a more realistic non-linear association regarding the climatic covariates and the health response based on evidences of the suscetibility of *Aedes Aegypti* specie to different weather conditions (Brunkard et al., 2008; Thomas et al., 2006; Githeko et al., 2000; Hii et al., 2009) (see Figure C.5). These shapes are still difficult to explain based on the theoretical evidences described in chapter 2 (Hii et al., 2009; Fuller et al., 2009; Johansson et al., 2009).

Consequently this third stage was conducted searching for the best fit model for predictive purposes of dengue cases in Zulia state with an accurate understanding on the shapes of the smoothed functions. In addition the necessity of seeking levels of temperature, rainfall and the El Nino3.4 at which the incidence rates sharply rise led to a comprehensive analysis at this stage. Despite many of the considerations set out in this chapter are based on seeking the GAMM model which better adapts for predictive purposes of dengue fever in Zulia state. The proposed models from all different stages are considered as an exploratory analysis aimed to that goal.

According to some authors a model for predictive purposes should avoid the effects observed over a long period of time (Rodier et al., 2004). Consequently a new approach was conducted at this stage by removing the smoothed term for the long term trend included into the previous stage. In addition, the peaks observed in Figure C.5 during the years 2002 and 2007 via the estimated smoothed curves for the long term trend of *timeserie*, might justify the construction of a new GAMM structure to clarify the reasons for such a pattern.

In this respect by removing the smoothed term of the long term trend, the replacement by a two-way interaction term included as a tensor product in a pure additive way (Underwood, 2009; Wood, 2005) might be an option to explain the pattern observed at a long term. Thereby the interaction terms could include all possible combinations using the climatic covariates. However, only one interaction term was included based on maximum temperature index and mean rainfall at 1 month delay respectively. This decision was made because it was noticed that changes in the El Nino3.4 values did not affect the predictive values. Thereby keeping in mind the predictive purpose, it was decided to arrange a new dataset monthly aggregated from January 2002 to June 2009 due to the availability of the dataset. And a subsequent validation of the predictions was proposed using data from July 2009 to June 2010.

Under such circumstances new options were explored in searching for the best

GAMM model for predictive purposes. Thus the three new options proposed are specified in detailed below.

First option: This included smoothed terms for mean rainfall and maximum temperature at 1 month delay respectively. The Nino3.4 with 6 months lagged effects was also included, plus a two-way interaction term combining maximum temperature and mean rainfall. In addition the non-climatic covariates were implemented in a categorical form previously specified at the second stage. An smoothed term for the variable *month* was also added within the structure accounting for a seasonal trend. However the smoothed long term trend was removed from the structure.

Second option: Unlike the previous one, the interaction term was split and only main effects for climatic covariates were assessed in conjunction with the non-climatic covariates in a categorical form. The smoothed term for the seasonal trend was included. But smoothed long term trend was removed from the structure.

Third option: This option was derived from the first one. This included main effects for maximum temperature and mean rainfall at 1 month delay, plus a two way interaction term between temperature and rainfall indexes. However the El Nino3.4 at 6 months delay was treated in a categorical form with three levels covering a range from 24.7° C to 28.43° C. The non-climatic covariates were also considered in a categorical form. In addition the smoothed term for the seasonal trend was included and smoothed long term trend removed.

5.5 Analysis of results

As it was previously specified, the construction of the optimal spatio-temporal GAMM structure for predictive purpose of dengue fever in Zulia state was built following a systematic approach due to the reduced number of technical reports in this field, that

could guide a potential solution adapted to the real dataset. To this end the construction of the optimal model was based on three different stages. This approach was conducted to investigate potential patterns that could contribute with the final version for predictive purposes. Even though each modelling stage was defined following its own specification, they complement each other by reaching the last version described on the third stage.

The first stage was conducted to investigate the optimal delay time and the fitting climatic covariates that best explain the dynamics on dengue transmission in Zulia state from January 2002 to December 2008. To this end a set of Negative Binomial GAM's models (Wood, 2005) were assessed including a smoothed term of temperature (amongst the options of: maximum and average) at 0,1 and 2 months delay respectively. Likewise a smoothed term of rainfall (amongst the options of: maximum, average and minimum) at 0, 1 and 2 months delay. And a smoothed term of the El Nino3.4 with 6 months lagged effect was added within the structure. The non-climatic covariates of altitude, percentage of households with access to the piped water supply and area expressed in square kilometers were also treated as smoothed terms in a continuous form.

After a forward selection with lowest Akaike Information Criterion (AIC), the findings revealed that maximum temperature and mean rainfall at 1 month delay better captured the dynamic of dengue transmission in Zulia state. However a new selection process was performed, aimed to identify the optimal degree of freedom for the smoothed terms included within the Negative Binomial GAM's structures. To this end free restriction on the degree of freedom of smoothed functions was compared with fixed parameter at 5 and 3 respectively. As a result the AIC levels demonstrate that they increase as the degree of freedom decreases.

The findings at this stage revealed that the structure with free restriction on the degree of freedom was the best option in detecting the optimal lagged effects of climatic

covariates associated with the epidemiological dataset from Zulia state. In addition, all the smoothed terms of: maximum temperature at 1 month delay, average rainfall at 1 month delay, the El Nino3.4 at 6 months delay, in conjunction with altitude, percentage of households with piped water supply and area, were found highly significant (see Table 5.2) reaching an explained deviance of 36.5%.

However, it was seen a clear evidence related to overfitting (Wood, 2006) because of the free restriction on the degree of the freedom. This is a matter that needs to be overcome in subsequent analysis.

Parametric coefficients		
Covariate	Estimate	p-value
(Intercept)	-9.3970	$< 2e - 16$
regionCosta Oriental	0.6648	0.00172
regionGuajira	0.2759	0.75708
regionPerija	-0.9784	0.05853
regionSur del Lago	1.3109	0.01234
Smooth terms		
Covariate	edf	p-value
s(maxtempclag1)	5.345	$< 2e - 16$
s(meanrainfratelag1mm)	8.289	$< 2e - 1$
s(nino3.4lag6)	8.895	$< 2e - 16$
s(altitude)	3.502	0.000467
s(perhwater)	8.711	$< 2e - 16$
s(area)	3.261	$5.29e - 11$

Table 5.2: Outcomes of the GAM structure at the first stage: optimal lagged effects

In addition, some visual evidence regarding the smoothed functions implemented with free restriction in the degree of freedom (coded in *mgcv* package as $k=-1$) revealed that incidence rates were affected in a non-linear fashion by the climatic covariates (Figure 5.1). However although it seems to be a good approach, the smoothed curves revealed distorted shapes regarding the climatic covariates that cannot be explained by the theoretical considerations described in chapter 2. Even more this approach needs to be fixed because it may cause a conflict of overfitting (Wood, 2006).

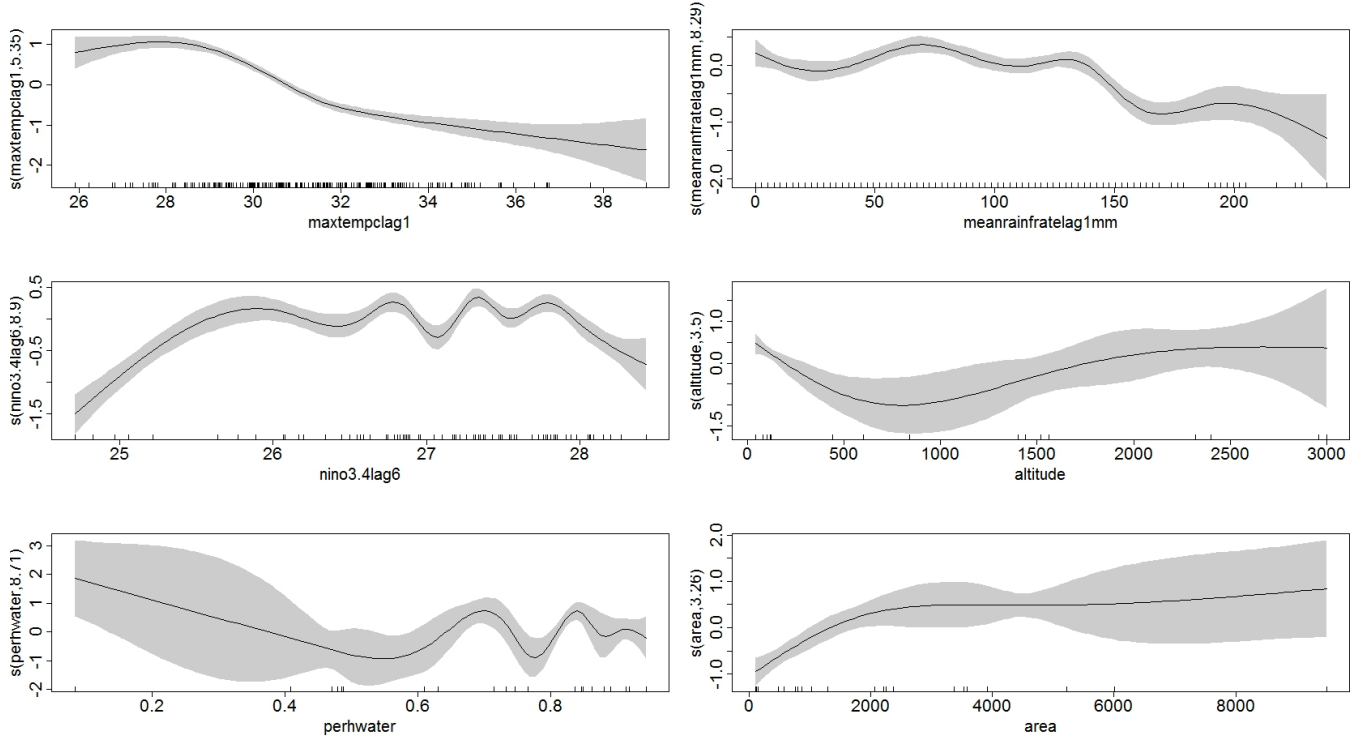


Figure 5.1: GAM-smooth curves with 95% confidence intervals from the optimal structure at the first stage

Furthermore, it is seen in Figure 5.1 that the non-climatic factors of: altitude, area (expressed in square kilometers) and percentage of households with access to the piped water supply network, treated as smoothed terms, might not be the best option due to a gap observed amongst the X-axis measures. Thereby these patterns might be understood as inaccuracy on the function's shapes. Under such circumstances a more accurate spatio-temporal modelling of dengue fever in Zulia state needs to be achieved.

However before moving into a more accurate spatio-temporal structure a validation process was performed to assess the statistical assumptions of independence and homogeneity of the residuals (Zuur et al., 2009; Wood, 2006) (see Appendix C). To this end the Pearson and Deviance residuals were plotted against: fitted values, explanatory variables and region units. Thereby the validation process was based on a simple visualization of patterns along the residual plots (Zuur et al., 2009; Wood,

2006). Although, some authors state a similarity of outcomes when the Pearson and Deviance residual for count data is conducted (Zuur et al., 2009; Wood, 2006). The analysis implemented in this study included both approaches for comparative purposes.

Consequently the residuals against the fitted values were used to test for homogeneity in conjunction with the residuals against the covariates. The autocorrelation functions (ACF) and the partial autocorrelation functions (PACF) at regional level were also performed to appraise the independence assumption over time (Zuur et al., 2009) (see Appendix C).

These graphs revealed a slight pattern in most of the residual plots against the fitted values but not serious enough to be concerned. However a marked temporal autocorrelation with an exponential decay over time was observed along the ACFs and PACFs plots (see Appendix C). These findings need to be confirmed with a more accurate spatio-temporal structure.

Furthermore, the inclusion of municipalities instead of regions was a better practice based on evidences of homogeneity from the residual plots across the five main regions that comprise Zulia state (see Appendix C).

Towards the construction of a more accurate spatio-temporal modelling of dengue fever in Zulia state a new structure was built at the second stage based on the limitations observed from the previous one. The structure was based on a GAMM model including municipality as random effect. In addition the non-climatic covariates of: altitude and percentage of households with access to the piped water supply network were linearly assumed in a categorical form. This approach also included smoothed term to capture the seasonal trend represented by a cyclic argument using the R-package *mgcv* (Wood, 2006). A smoothed term of *timeserie* was also added to measure the long term trend (Staelens et al., 2012; Underwood, 2009). However this analysis might only allow the exploration of new outcomes that could benefit the construction of the predictive

approach.

As a result, the findings at this stage revealed that all smoothed terms were statistically significant at a level below 5% . And these outcomes were explained with a 68.8% of accuracy (see Table 5.3).

Parametric coefficients		
Covariate	Estimate	p-value
(Intercept)	-8.65	$< 2e - 16$
altgrp2	-0.39	0.06
perhwatergrp2	-0.38	0.082
Smooth terms		
Covariate	edf	p-value
s(municipality)	14.6	$2.17e - 11$
s(timeserie)	8.17	$< 2e - 16$
s(month)	7.20	$< 2e - 16$
s(maxtempclag1)	3.88	$3.8e - 15$
s(meanrainfratelag1mm)	1.36	0.0135
s(nino3.4lag6)	2.63	$3.63e - 07$

Table 5.3: Outcomes of the GAMM using municipality as a random effect from the second stage

Moreover the smoothed functions of rainfall, temperature and the El Nino3.4 followed non-linear shapes in a more realistic way (see Figure C.5). However they are still difficult to explain based on theoretical evidence described in chapter 2 (Hii et al., 2009; Johansson et al., 2009; Fuller et al., 2009).

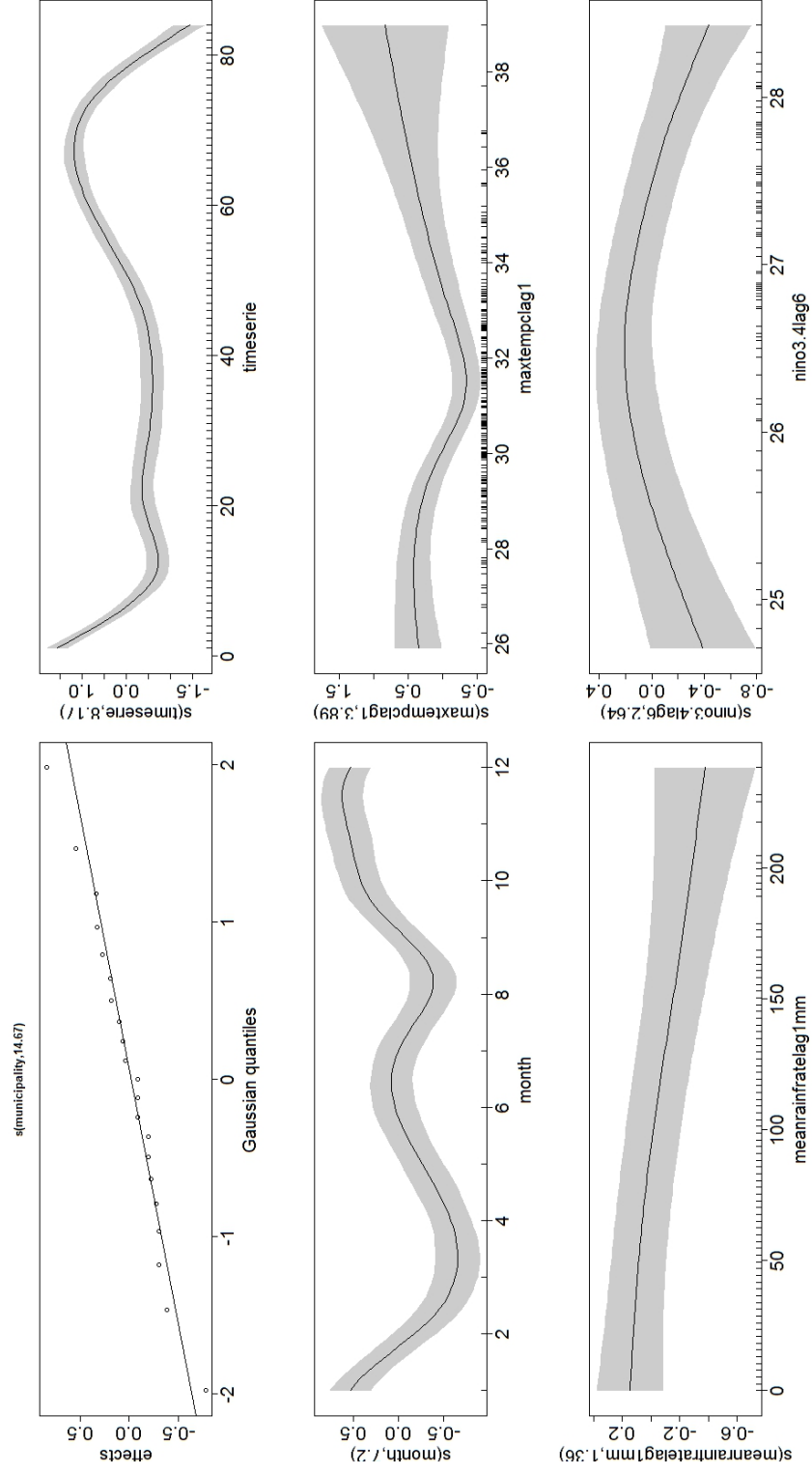


Figure 5.2: GAMM estimated functions using random effect at municipality level, with 95% confidence interval from the second stage

A validation process at this stage was also performed to assess the statistical assumptions. Slight patterns in most of the residual plots against the fitted values were observed (see Appendix C) but not serious enough to be of concern. In addition a marked temporal autocorrelation with an exponential decay over time was also confirmed along the ACFs and PACFs plots. Consequently it was suggested that an autoregressive (AR) structure of order 1 could be the best temporal structure to be included for future extensions of the GAMM model (Root, 2011).

This validation process cannot be completed if the Moran's I on the residuals is not performed to measure the spatial autocorrelation amongst neighbouring municipalities (Pfeiffer et al., 2008). As a result the GAMM structure using municipality as a random effect in conjunction with non-climatic covariates in a categorical form and smoothed terms for: seasonal trend, long term trend and the climatic covariates were tested with the Moran's I. Thereby the R-code was adapted including a matrix of spatial proximity based on two different approaches for comparative purposes (see Appendix C). These approaches included Simple Contiguity Neighbours and Queen Contiguity Neighbours, having common borders and corners respectively (Stevenson, 2011; Bivand, 2007).

As a result, both approaches revealed that the null hypothesis was accepted at 5% level, meaning a random spatial pattern across the 21 municipalities which comprises Zulia state over January 2002 to December 2008. In addition, a geographical visualization of the deviance residuals at municipal level confirms a random spatial pattern across the units (see Figure 5.3).

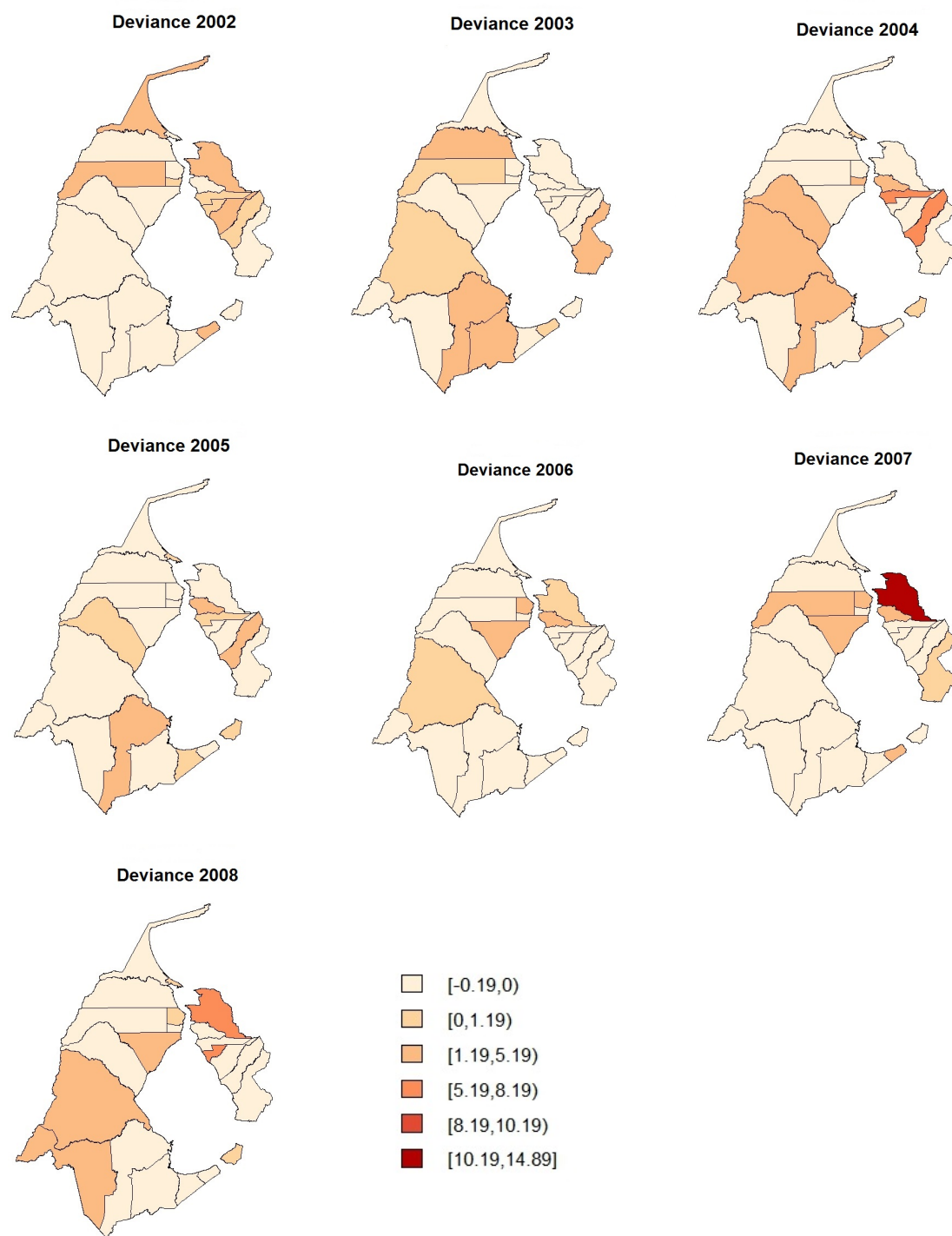


Figure 5.3: Mapping Moran's I: Random spatial autocorrelation from 2002 to 2008 from GAMM structure at the second stage

Because the main goal of this research is the development of a predictive model of dengue cases in Zulia state with a high standard of accuracy, a third stage on modelling needs to be conducted for this purpose.

As was previously discussed in the model specification, the third stage was conducted searching for the best fit model for predictive purposes of dengue cases in Zulia state with an accurate understanding on the shapes of the estimated smoothed functions. It was also emphasized that smoothed terms for the long trend need to be removed for predictive purposes (Rodier et al., 2004).

As a result, three new options were tested to find the best fit model (see the specification in section 5.4). The model specifications are summarized as follows:

First option: Smoothed terms for mean rainfall and maximum temperature at 1 month delay respectively. Smoothed term for the El Nino3.4 at 6 months lagged effects. Smoothed term for seasonal trend. Two-way interaction term between maximum temperature and mean rainfall. The non-climatic covariates were treated in a categorical form and smoothed long term trend was removed.

Second option: Smoothed terms for climatic covariates used as main effects in conjunction with the non-climatic covariates in a categorical form. Seasonal trend smoothed term included and the long term trend removed.

Third option: This option included smoothed term of climatic covariates of temperature and rainfall, plus a two-way interaction term between temperature and rainfall. Non-climatic covariates and the El Nino3.4 were treated as categorical variables. Also smoothed term for seasonal trend was included.

Although a model selection process was performed based on the lowest Akaike Information Criterion (AIC). This theoretical outcome can not provide an accurate

understanding of the real shapes of the smoothed functions. As a result, a more pragmatic decision was made based on a realistic shape of the smoothed functions involved within the model. Under such circumstance, the second option was identified as the best choice. The selected model includes smoothed terms with main effects of temperature, rainfall and the El Nino3.4. And the simpler version was used with fixed degree of freedom in 2.

A summary of outcomes from the final selected model for predictive purposes is shown in Table 5.4. And the estimated curves are displayed in Figure 5.4.

Parametric coefficients		
Covariate	Estimate	p-value
(Intercept)	-8.55	$< 2e - 16$
perhwatergrp2	-0.42	0.05
Smooth terms		
Covariate	edf	p-value
s(municipality)	13.57	0.0004
s(month)	5.64	$< 2e - 16$
s(maxtempclag1)	1.98	$< 2e - 16$
s(meanrainfratelag1mm)	1.92	$< 2e - 16$
s(nino3.4lag6)	1.97	$2.8e - 07$

Table 5.4: Outcomes from the final GAMM structure for predictive purposes

These findings confirm that all the covariates were statistically significant to explain the dynamics of dengue fever in Zulia state. And the smooth curves from Figure 5.4 fairly capture dengue transmission regarding the weather conditions.

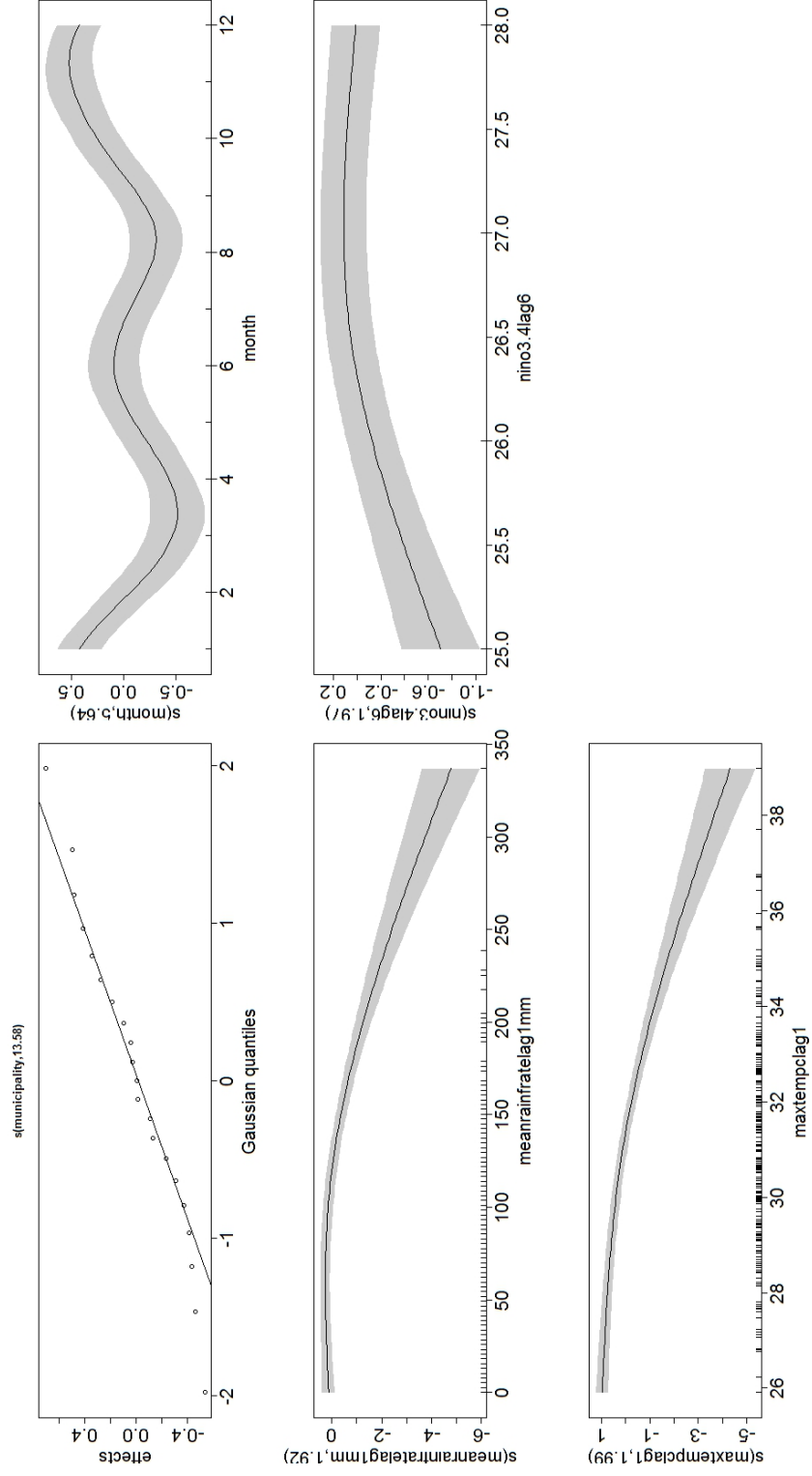


Figure 5.4: Estimated smooth curves from the final GAMM structure for predictive purposes

The percentage of households with access to the piped water supply network covering from 81 to 100%, showed to be negative significant, when comparing with the baseline group from 0 to 50% of the households with access to the public service.

In addition it was found not evidence on the variation of the disease amongst levels of altitude. This finding confirms that heights are not longer relevant to explain the dynamics of dengue transmission. In this context recent studies have found new outbreaks of dengue fever in regions located on high mountains (Epstein et al., 1998). A particular interest the outbreaks in countries of the South American cone (Carbajo et al., 2012; Gurtler et al., 2009).

On one hand, a visual representation of the smoothed functions for climatic covariates (see Figure 5.4) confirm the biological principles regarding the susceptibility of *Aedes Aegypti* to different weather conditions (see chapter 2). But on the other hand, the levels at which the incidence rates sharply rise to a critical public health level were clearly identified.

At this respect, critical level of rainfall was found between 0 to 100 mm. It confirmed previous statements in which low levels of rainfall facilitates the breeding sites of the mosquito population owing perhaps to the use of artificial water containers during the dry season (Bohra and Andrianasolo, 2001). In contrast it is seen in Figure 5.4 that heavy rainfall (greater than 100 mm) sharply diminish the incidence rates of dengue fever. These latest findings were also confirmed by some authors who state that heavy rainfall may destroy the larvae of *Aedes aegypti* mosquitoes (Hii et al., 2009). In addition the non-linear association regarding rainfall and health outcome was evident in Figure 5.4 confirming previous statements (Hii et al., 2012; Hu et al., 2010) found in different parts of the world as findings from the previous analysis.

The impact of temperature on the transmission of dengue fever in Zulia state was also found in a non-linear fashion. This finding confirms previous studies conducted

in different parts of the world (Brunkard et al., 2008; Thomas et al., 2006; Githeko et al., 2000). The outcomes using a dataset from Zulia state revealed that the health response was at risk when temperature fluctuates between 26°C to 30°C, but this risk sharply drops when temperatures exceeds the 30°C.

In regard with the El Nino3.4, although no evidence was found in previous studies regarding the critical levels of the El Nino3.4, the outcomes from this research identified for the first time that risk of the El Nino3.4 fluctuates between 26.5°C to 28.0°C. Being perhaps a reference point to be discusses further. For more detail the R-code of the final selected GAMM model for predictive purposes was listed in Appendix C.

These promising results were used to predict cases of dengue in Zulia state in the subsequent analysis described in section 5.6.

5.6 Predicting dengue cases in Zulia state

The use of a satellite remote sensing dataset in the scientific field has greatly improved the construction of predictive models for the outbreak of infectious diseases (Connor et al., 1998; Ceccato et al., 2007). But although various studies have been aimed to an early detection of malaria (Bouma and Van der Kaay, 1996; Bouma et al., 1997; Connor et al., 1998; Thomson et al., 2004; Abeku et al., 2004) only few researches have been designed for the outbreaks of dengue fever in different parts of the world (Lowe et al., 2013; Hii et al., 2012; Descloux et al., 2012; Hu et al., 2010).

Predicting outbreaks of dengue fever has been recently identified as the main global goal to be achieved in the near future, according to the WHO (Chan, 2012). This international body has also acknowledged its valuable use towards the implementation of effective interventions for control and prevention of the disease, aimed to reduce mortality and morbidity on human population.

Although some models to predict imminent epidemics of dengue fever have dominated the field using time series with a Poisson distribution for count data (Hii et al., 2012; Descoux et al., 2012; Hu et al., 2010), they have been climate-based structures following a non-linear fashion in the mentioned cases. However a similar study has been recently proposed in Brazil (Lowe et al., 2013) using a Generalized Linear Mixed model (GLMM). This approach although implemented a more diverse set of covariates including not only climatic factors but also a set of non-climatic variables. They were arranged following a linear assumption.

What is being proposed in the current study is a new approach for predictive purposes of dengue cases, based on the Generalized Additive framework. This included not only a non-linear association between climatic factors and response variable but also a set of non-climatic factors implemented in a categorical form which might trigger a more robust prediction (Rodier et al., 2004). Thereby, this approach could bring together the advantages from the previous methodologies.

Even more, a comprehensive assessment of prediction accuracy was achieved using an out-of model dataset (Rodier et al., 2004; Hii et al., 2012; Lowe et al., 2013). For this end, the final GAMM structure was run with data from January 2002 to June 2009 and the validation of the prediction was achieved with data from July 2009 to June 2010 monthly aggregated at municipality level.

In this context, experts have suggested a period of time of five weeks in advance of an epidemic as optimal time for launching of preventive actions (Connolly, 2005). However the findings from this study found a remarkable time of four months in advance for the prediction of dengue cases in Zulia state. This finding was also validated with a 90% predictive interval. The months included within the good prediction were: July, August, September and October from 2009. However, these outcomes constitute a good achievement, which led to the conclusions expressed in the previous statements. More

analysis is needed to investigate the real pattern between the predicted dengue cases and the actual observed cases by timely addressing different period of time within the validation process. As a result these findings might be sensitive to the period of time chosen for the validation. Predictive models suggest a period of time of five weeks for detecting epidemics in advance (Connolly, 2005), consequently more analysis is needed to clarify this matter.

Figure 5.5, illustrates the prediction values using the new dataset over the period July 2009 to June 2010 monthly aggregated. The dashed line shows the predictive values, while the black points represent the observed cases of dengue in the whole of Zulia state. The statement regarding closeness in predictive values over the four consecutive months of: July, August, September and October can be seen in this graph. However a more accurate validation was achieved later.

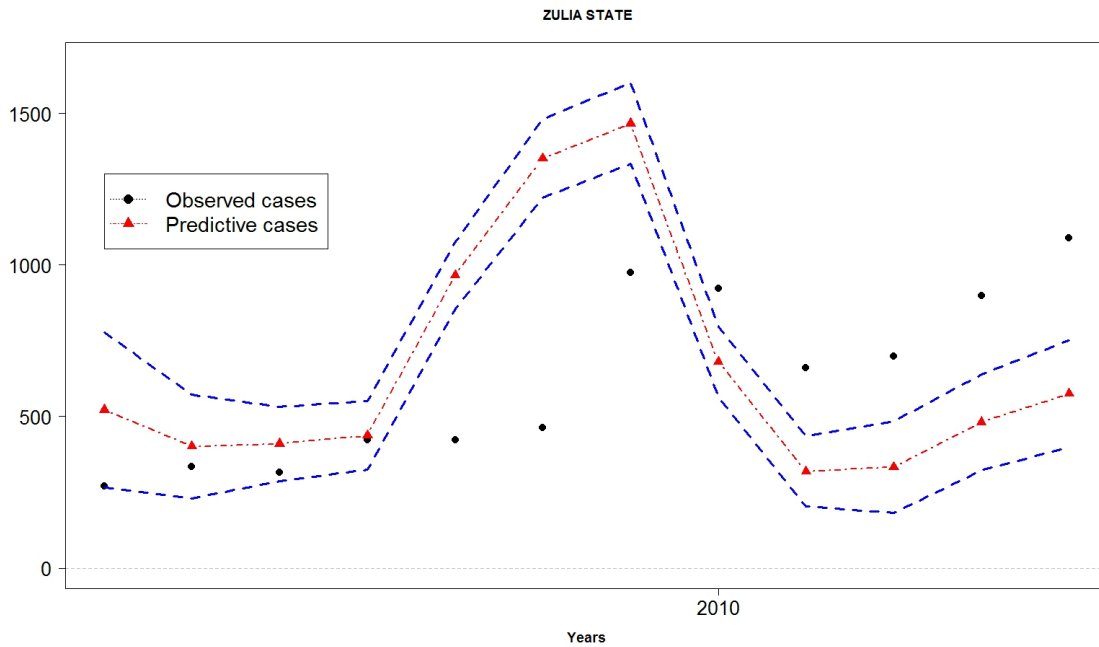


Figure 5.5: Predictive trend line from the final GAMM with 90% predictive interval using data from July 2009 to June 2010.

Therefore, the construction of predictive intervals for binomial distribution was per-

formed (Wang, 2010; Wayne, 1982) to confirm previous statement regarding the closeness in predictive values during the four consecutive months of: July, August, September, October from 2009. This technique was originally designed by Wayne (Wayne, 1982), and has demonstrated its performance in a wide range of applications within the medical field (Wang, 2010). In addition, the fact that count cases of disease usually follow a binomial distribution has enormously increased the versatility of this technique (Wayne, 1982; Wang, 2010).

As a result, the estimated predictive intervals based on the future dengue number of success were listed in Table 5.5. It can be seen that observed cases of dengue over: July, August, September and October had values of: 269; 334; 315 and 423 respectively. And these values fall into their respective predictive intervals at 90% level of confidence. Unlike, during the remaining months the observed values differ considerably from their predictive intervals.

Month	Predictive interval
July	$522 \pm 255.7 = [266.2; 777.7]$
August	$401 \pm 170.3 = [230.6; 571.3]$
September	$410 \pm 122.4 = [287.5; 532.4]$
October	$437 \pm 113.0 = [323.9; 550.0]$
November	$966 \pm 11.6 = [854.4; 1077.6]$
December	$1351 \pm 129.0 = [1221.9; 1480.0]$
January	$1466 \pm 133.1 = [1332.8; 1599.1]$
February	$680 \pm 116.6 = [563.3; 796.6]$
March	$319 \pm 115.4 = [203.6; 434.4]$
April	$333 \pm 150.5 = [182.4; 483.6]$
May	$481 \pm 157.4 = [323.5; 638.4]$
June	$575 \pm 176.0 = [398.9; 751.0]$

Table 5.5: Binomial predictive interval based on the final GAMM structure, at 90% CI, using data from July 2009 to June 2010

Having achieved these promising results the next step need to be the construction of an Early Warning System (EWS) of dengue fever in Zulia state. According to the protocol about the development of an EWS of infectious disease discussed in detail in

chapter 2, the construction of a forecast model based on the predictions of explanatory variables (Hii et al., 2012; Lowe et al., 2013) need to be considered. At this respect the next goal to be achieved on further studies is the development of a forecast model for explanatory covariates, including the climatic factors. In addition, because historic patterns of count dengue cases have a tendency to recur over time, a Bayesian framework has been pointed to be the optimal approach because previous cases can be used for modelling the posterior counts of dengue fever (Lowe et al., 2013).

5.7 Discussion and conclusions

Clear evidence was provided by the exploratory analysis conducted in chapter 3, confirms a complex pattern in modelling the real dynamics on dengue transmission by using epidemiological data of dengue fever from Zulia state, Venezuela. A similar study undertaken in Brazil (Lowe et al., 2011) revealed a technical limitation in modelling such relationship due to the linear assumption from the GLMM approach implemented.

In this context, although non-linear structures were confirmed in some other studies carried out in different parts of the world (Hii et al., 2009; Johansson et al., 2009; Fuller et al., 2009). A more flexible approach was developed in this chapter via a semi-parametric structure aimed to extend the GLMM model. Thereby the Generalized Additive framework (Hastie and Tibshirani, 1986) was implemented for the first time with the inclusion of unknown smoothed functions that better captured the complexity of the associations amongst the covariates and the response variable tailored to explain the real dynamics on dengue transmission.

As a result the construction of a spatio-temporal modelling via the novel structure of Generalized Additive framework (Wood, 2006) for predictive purposes was conducted in a systematic way. This approach was implemented to illustrate the evolution in building a spatio-temporal GAMM model for predictive purposes of dengue fever in

Zulia state. The process of construction relied on three different stages. Each with its own model specification but they complemented each other towards the final version.

Formal model selection process based on the lowest Akaike Information Criterion (AIC) in conjunction to some intuitive common sense, were ultimately the decision-making procedures for the selection of the optimal models. A set of Negative Binomial GAMs models were performed at the first stage to investigate the optimal lagged effect and the climatic covariates associated with the incidence rates of dengue fever. To this end the GAMs structures were implemented using region as a factor. And smoothed terms for non-climatic covariates were also included. A forward selection process was conducted via a lowest Akaike Information Criterion (AIC) at this stage. As a result clear evidence was revealed that maximum temperature and average rainfall correctly captured the dynamic of dengue transmission in Zulia state at 1 month delay. In conjunction to the El Nino3.4 at 6 months.

The validation process was performed based on a simple visualization of the residual plots (Zuur et al., 2009; Wood, 2006). As a result a marked temporal autocorrelation with an exponential decay over time was noticeable at this stage.

The second stage, allowed a more comprehensive exploratory analysis in modelling the dynamics of dengue transmission. Some limitations from the previous stage led to a more accurate spatio-temporal modelling of dengue fever in Zulia. At this stage a spatio-temporal GAMM model based on random effect for geographical entities of municipalities was conducted. And additional smoothed terms of seasonal and long term trend were included aimed to explore the outcomes that could benefit the construction of a predictive structure. Furthermore non-climatic covariates of altitude and percentage of households with access to the piped water supply network were implemented in a categorical form. An argument gamma defined equal to 1.4 ($\gamma \approx 1.4$) was also included to avoid overfitting issues (Wood, 2006). And finally a fixed term of $k=5$ was implemented for a better performance of the smoothed terms.

As a result the findings revealed that all smoothed terms were statistically significant at 5% level and the outcomes were explained with 68.8% of accuracy. However the smoothed functions for climatic factors were still difficult to understand based on biological evidences regarding the susceptibility of the species to weather conditions (discussed in detail in chapter 2).

Although a marked temporal autocorrelation was also noticeable at this stage, not evidence of spatial autocorrelation was found via the Moran's I test.

Following the systematic approach the third stage was built searching for the best fit model adapted to predictive purposes, with an accurate understanding of the smoothed functions for climatic covariates. Aimed to this goal three different options were tested by removing the smoothed term of the long term trend and the replacement with some other structures such as: interaction terms and extra categorical forms. As a result a more pragmatical decision was made based on a realistic shape of the smoothed functions. The selected model accurately explained the smoothed shapes based on the biological principles described in chapter 2. Moreover the levels at which the incidence rates sharply rise to a critical level were clearly identified. Levels at risk of mean rainfall for instance, were located below 100mm, confirming previous studies undertaken in other part of the world (Bohra and Andrianasolo, 2001). According to these authors low levels of rainfall facilitates the breeding sites of the mosquito population perhaps because it promotes the use of artificial water containers during a dry season. The critical levels of temperature regarding dengue transmission in Zulia state fluctuates between 26°C to 30°C, but it sharply drops when temperature exceeds the 30°C. This pattern has been also confirmed in previous studies conducted in different parts of the world (Brunkard et al., 2008). According to these authors, the incubation period of the virus drops to seven days when temperature fluctuates between 32°C to 35°C and consequently it affects the speed on dengue transmission.

Finally although no relationship was revealed in previous studies regarding the critical levels of the El Nino3.4. The outcomes from this research identified that the risk on the El Nino3.4 fluctuates between 26.5°C and 28.0°C being a reference point to be discussed further. In addition, the final GAMM structure was run with monthly aggregated data from January 2002 to June 2009 for a better performance to trigger the predictions.

Two different methodologies have recently been implemented for predictive purposes of dengue fever in different parts of the world. Some authors have used Time Series based on climatic factors associated with dengue transmission in a non-linear fashion (Hii et al., 2012; Descoux et al., 2012; Hu et al., 2010). However some others prefer the inclusion of non-climatic covariates in conjunction with the climatic ones following a linear assumption (Lowe et al., 2013).

However as a result of the approach undertaken in the case study of Zulia state, an optimal prediction was obtained, validated at four months in advance with 90% predictive interval (Wang, 2010; Wayne, 1982). This was based on a data-out of model (Rodier et al., 2004; Hii et al., 2012; Lowe et al., 2013) from July 2009 to June 2010. These promising results will lead the construction of an Early Warning System (EWS) of dengue fever in Zulia state, Venezuela as the next goal to be accomplished in further work. To this end, a forecast model based on the prediction of the explanatory variables (Hii et al., 2012; Lowe et al., 2013) need to be obtained. And the implementation of a Bayesian framework might allow the inclusion of historic patterns of dengue counts to drive the posterior predictions (Lowe et al., 2013).

Finally, a clear temporal autocorrelation revealed in the previous stages of modelling may lead to carefully propose an extension of the GAMM by including a temporal correlation structure.

Chapter 6

General discussion and conclusion

6.1 Introduction

It was the potential of academic research to have a social impact that led me to propose this dissertation framed by the real needs of the public health sector in Venezuela. In 2008, I was infected by dengue fever in my country (Venezuela) but thankfully did not suffer any serious complication. However according to statistics released by the WHO a large number of people across the globe do not enjoy the same outcome.

As a result the Spatial Epidemiology of Tropical diseases linked to the environmental conditions, became my main research interest. This is particularly a fascinating area of investigation because its relevance to real world needs, and because each dataset potentially offers endless alternatives, from which the optimal approach needs to be selected. The very great challenge of finding the best model that suits the complexity of the dataset became the most enjoyable experience during this journey.

According to the WHO statistical modelling has been acknowledged as the most effective mechanism for driving public health response to combat and prevent infectious diseases on a global scale. Consequently this research may hopefully contributes to im-

prove public health in Venezuela, and perhaps can be extended to other geographical contexts.

6.2 Discussion and conclusions

The overall aims from this thesis are summarized as follows:

- To develop a Spatio-Temporal modelling of dengue fever in Zulia state from 2002 to 2008, via a generalized additive model with random effects.
- Use this model for predictive purposes towards the construction of an Early Warning System (EWS) of dengue fever in Zulia state.

Addressing all of these goals, led me to develop a systematic approach to better justify the statistical tools to be used, which perfectly adapted to the complexity and diversity of the sources of information available.

First of all a comprehensive exploratory analysis was conducted to underline the most salient features in the dynamics of dengue transmission in Zulia state, between 2002 and 2008. As a result a visual representation of the spatial and temporal variation of dengue incidence revealed a wide heterogeneity in rates during 2002, 2006 and 2007 across the 21 municipalities that comprises Zulia state. However the highest variation was observed for 2007 and led to an exploration of potential factors associated with the distribution of the disease at spatial and temporal scale. As a result a range of categories were examined based on the literature review described in chapter 2. These categories included a set of climatic, geographic, demographic, socioeconomic and individual factors aggregated by weeks, months or year and at municipal, regional or at the whole of Zulia state.

Some of the findings at this stage revealed that the group at most risk was children between 5 and 14 years. The second and third most affected groups were those aged 15

to 24 years and 0 to 4 years respectively. However the trend related to gender showed a similar pattern for both male and female groups. At this respect, more advanced techniques were developed in this thesis to validate these findings.

In regard to climatic factors, the initial information about local weather conditions in Zulia state was provided by the local meteorological stations. Moreover this project benefited from remotely sensed climatic data provided by the international agency of the National Aeronautics and Space Administration's (NASA) because the absence of available continuous records from the local stations.

Previous studies have confirmed the usefulness of satellite remote sensing data of climatic information in epidemiological studies (Zacarias and Andersson, 2011; Fuller et al., 2009; Lowe et al., 2011). The availability of real-time and accurate information at a low cost (Connor et al., 1998) is the main reason of such decision. As a result the exploratory analysis conducted in Zulia state included: average, maximum and minimum temperature ($^{\circ}\text{C}$) monthly aggregated from January 2002 to July 2010. In conjunction with average, maximum rainfall rate (mm) and the El Nino3.4 index. This information was allocated into four symmetric and non-overlapping grids of dimensions 2.0° latitude by 2.0° longitude covering the whole Zulia state.

Another important issue that was treated in this thesis was the empirical estimation of optimal delay time between the climatic factors and the incidence rates of the disease (Rothman et al., 2008). According to the exploratory analysis the potential delay time for temperature and rainfall indexes were: 0, 1 and 2 months in conjunction with 6 months delay for the El Nino3.4 index.

It was observed through the exploratory analysis that average temperature in Zulia state fluctuates between 22.9°C to 29.7°C providing an ideal condition for the development of *Aedes aegypti* species (Hii et al., 2009) and consequently the spread of the disease. It was observed a seasonal pattern regarding the hot conditions via the evo-

lution of average temperature. The hottest temperatures were markedly noticeable during February to April and from July to September. However maximum and minimum temperature were also examined seeking for greater variability across the grids which could better justify a spatio-temporal analysis of dengue fever in Zulia state.

It was seen as well that maximum temperature was widely spread across the geographical grids which might justify the inclusion of this index for modelling purposes. In addition a scatterplot of monthly aggregated maximum temperature at 1 month delay averaged per grid versus count cases of dengue on the whole Zulia (covering 84 months) from January 2002 to December 2008 was performed. But although it was difficult to reach a conclusion regarding the real shape of such association, based merely on this plot, the graphs suggested a severity of the disease by reaching a peak of 3,756 cases when the maximum temperature was around 30°C. By contrast, when the temperature was greater than 31°C the cases of dengue drastically diminished. However this complex association needs to be confirmed with more advanced techniques in subsequent analysis.

Moving on to the analysis of monthly rainfall rate aggregated per grid in the whole of Zulia, a time series plot showed the evolution of average rainfall from 2002 to 2008. It revealed a range from 0 to 284 mm, and a seasonal pattern with two major periods of intense rainfall: the first from May to June and the second from October to November. In addition a plot was made to assess the potential relationship amongst average rainfall and cases of dengue fever at 1 month delay. Although it was difficult to reach a conclusion about the real shape of such association using a scatterplot, it showed that levels of rainfall greater than 80 mm caused drastically diminution on dengue cases in Zulia state, whereas values between 0 to 80 mm caused dengue cases highly noticeable.

As stated above the most relevant finding from the exploratory analysis was the one related to evidences of a non-linear association between climatic factors that could better fit the complex dynamic regarding the outbreaks of dengue fever. Therefore a more

flexible approach for subsequent studies seems to be the Generalized Additive model (GAM) (Hastie and Tibshirani, 1984; Wood, 2006). This technique has emerged as a powerful tool because of its flexibility in substituting the linear structure by smooth functions which better model the real shape of the associations (Berg, 2005). Although this technique has been implemented in some areas such as finance (Berg, 2005), computer science (Lado et al., 2008), ecology (Murase et al., 2009; Guisan et al., 2002) and environmental studies (Terzi and Cengiz, 2009), it is to my knowledge the first time it has been applied to an epidemiological dataset of dengue fever for predictive purposes.

Further by following a systematic approach a second step was reached based on a Generalized Linear model approach (GLM) for count data (Pfeiffer et al., 2008; Lovett and Flowerdew, 1989; Schwartz, 1993; Zuur et al., 2009), designed to compare Relative Risk (RR) by exposure groups of age and sex using the data from the whole of Zulia. As a result a GLM model for count data using a Negative Binomial (NB) approach, accounting for heterogeneity on the variance (Hoeff and Boveng, 2007; Zuur et al., 2009) was conducted via both a frequentist and a Hierarchical Bayesian framework. These findings confirmed the previous exploratory outcomes and were consistent amongst them.

The implementation of both perspectives, the frequentist and the Bayesian approach allowed the comparison of outcomes and methodologies (Ismaila et al., 2007), regarding the estimation on Relative Risks (RR) of dengue fever across sex and age categories: 0 to 4; 5 to 14; 15 to 24; 25 to 44 and over 45 years old in the whole of Zulia state. A technical point: the frequentist version was built using the open statistical package R version 2.15.2 (Zuur et al., 2009) and the Bayesian inference was based on WinBUGS package (Gimenez et al., 2009; Abellan and Best, 2008; Lunn et al., 2000; Spiegelhalter et al., 2003).

Although the original intention was the construction of a spatio-temporal model using exposure factors at municipality level. Limitations in the epidemiological data

available aggregated within these scales restricted the construction of such analysis. Under such circumstances, it was assumed that age structure per gender remained unchangeable at municipality level over the period under analysis. However this assumption might be the cause of a potential overdispersion issue.

Aimed to this end, the next analytical task was the construction of a spatio-temporal modelling of dengue fever in Zulia state, via a Generalized Additive Mixed model (GAMM) (Wood, 2006) using municipality as a random effect. As a result a set of climatic and non-climatic covariates were included within the model and spatial and temporal autocorrelations assessed to justify a more accurate structure. At first stage the optimal lagged effects were identified for maximum temperature and mean rainfall at 1 month delay, in conjunction with the El Nino3.4 at 6 months delay. After a systematic approach with appropriate model selection process based on the lowest Akaike Information Criterion (AIC) the second structure was reached including: climatic covariates with optimal lagged effects and non-climatic factors of: altitude and percentage of households with access to the piped water supply network, both of them treated as categorical variables. At this stage spatial and temporal autocorrelations were assessed. As a result a noticeable temporal autocorrelation was observed with random spatial dependencies amongst municipalities.

Finally the third stage was built searching for the best fit model for predictive purposes. Although a temporal structure observed in previous stages was not included into this structure. It might be a potential extension to be achieved in further studies. However the final GAMM model from this stage was the starting point for predicting outbreaks of dengue fever in Zulia state. In addition it was validated with a new dataset out-of-model from July 2009 to June 2010.

This is the first time that a GAMM structure has been used for the purpose of dengue fever prediction using an epidemiological dataset with climatic and non-climatic covariates. According to some experts this version might contribute to a better per-

formance of the predictions (Rodier et al., 2004). Although there are still some improvements that need to be done. Promising results have been achieved for predictive purposes, validated at four months in advance with 90% confidence interval (Wayne, 1982; Wang, 2010) using an out-of-model dataset (Rodier et al., 2004; Hii et al., 2012; Lowe et al., 2013).

Accordingly the next step needs to be the construction of an Early Warning System (EWS) of dengue fever in Zulia state, Venezuela, based on the development of a forecast model (Lowe et al., 2013; Hii et al., 2012; Rodier et al., 2004). The implementation of a Bayesian framework has been suggested as the most effective practice because historic patterns of dengue counts can be used to drive the posterior predictions (Lowe et al., 2013) for a forecast model.

6.3 Limitations

An initial set of data was collected by personal contact at the early stage of this project. However as the researcher developed the skill for handling spatial epidemiological studies this data was replaced in many cases by more accurate information. In this regard one of the urgent needs observed in the public health sector in Zulia state is the implementation of a comprehensive surveillance system with a high standard of temporal and spatial resolution.

In the current project a variety of data sources were used, which produced a discrepancy in terms of spatial and temporal scales which required to be unified. Some other practical issue faced in this study was the poor resolution regarding epidemiological data to smaller geographical areas. This problematic situation needs to be overcome by the authorities responsible of the surveillance information to allow the development of advanced epidemiological studies with scientific quality.

It was also noticed that manual records, still exist as main source of information, despite a great effort to automate them. In this regard a step forward is needed in the implementation of automated records of the information in the public health sector. But perhaps the most important is the possibility of an easy access of information for scientific purposes.

Facing this reality a challenge ahead in the public health sector of Zulia state might be the construction of a surveillance network of dengue fever across all the entities that comprise this state. This project would provide the opportunity to join efforts amongst scientific community and health agencies at a local and national levels aimed to consolidate a robust tool with high standard of resolution.

In addition the records of dengue fever in Zulia state need to be completed with entomological data, types of serotypes circulating in different periods of time and indexes related to: socioeconomic, demographic, geographic and also climatic data. Furthermore information linked to intervention programs need to be included.

The inclusion of climatic data from a satellite remote sensing would be ideal. Otherwise, cooperation agreements with international agencies which provide such information. Finally more scientific publications on dengue fever in Venezuela are needed to trigger studies and effective control and prevention programmes for the welfare of the population.

6.4 Future directions

Aimed to the construction of an Early Warning System (EWS) for outbreaks of dengue fever in Zulia state, some technical studies need to be undertaken following four major directions, which have been summarised in Figure 6.1.

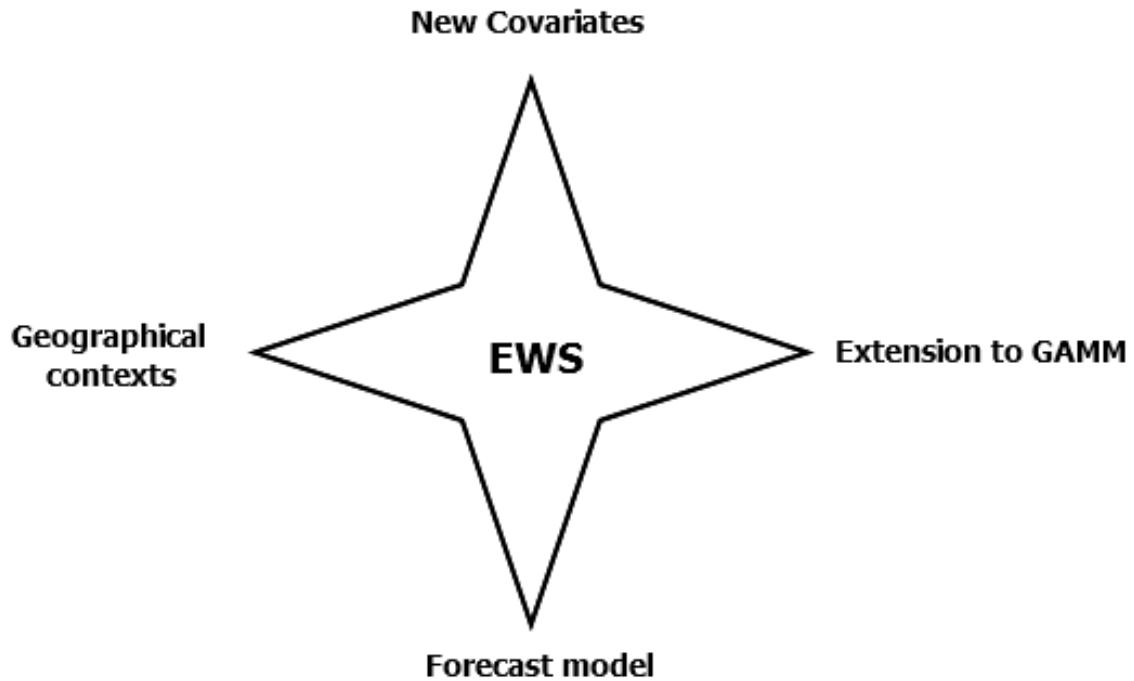


Figure 6.1: Directions for future work

New Covariates:

Some authors have emphasized the need to include entomological data into the studies of dengue transmission, arguing that the mosquito population can be easily eradicated at the very early stage of the larvae and pupae development (Garelli et al., 2009). As a result a spatio-temporal modelling for predictive purposes might provide a more accurate tool, if epidemiological data in conjunction to entomological information are included. Furthermore serotypes circulating, indexes quantifying the intervention programs in conjunction to human movements, could also be beneficial.

Geographical contexts:

According to the protocol for the construction of an Early Warning System (EWS) (see chapter 2), a predictive model needs to be validated in different geographical locations. In this context a recent Santander research travel grant awarded during the

last July 2013, allowed me to analyse the recent outbreak of dengue fever occurred in Argentina. The fact that this disease is spreading across Latin American countries by crossing borders reaching the Southern Cone, could be an opportunity to consider potential cooperative efforts between the scientific community and the public health sector at a local, national or multinational scale.

Forecast model:

A forecast model needs to be the next goal to be accomplished in further studies towards the construction of an Early Warning System (EWS) of dengue fever in Zulia state. To this end experts argue that the prediction of the explanatory variables need to be provided in some extent (Hii et al., 2012; Lowe et al., 2013). In addition some others state that the Bayesian framework might be the most effective option because it allows the inclusion of previous dengue counts to estimate the posterior prediction of count cases (Lowe et al., 2013).

Extension to GAMM:

Future extension of the GAMMs structure need to be implemented to account for temporal autocorrelation structure, based on the findings suggested in the validation process considered in this project. In addition improvements in the GAMMs structure are always potential studies to be achieved. Furthermore comparisons between the performance of the predictions between the linear and semiparametric structures, might be an interesting point of analysis.

Appendix A

Exploratory analyses

A.1 Exploratory analysis of socioeconomic factors

Scatterplot - incidence rate and non-climatic factors

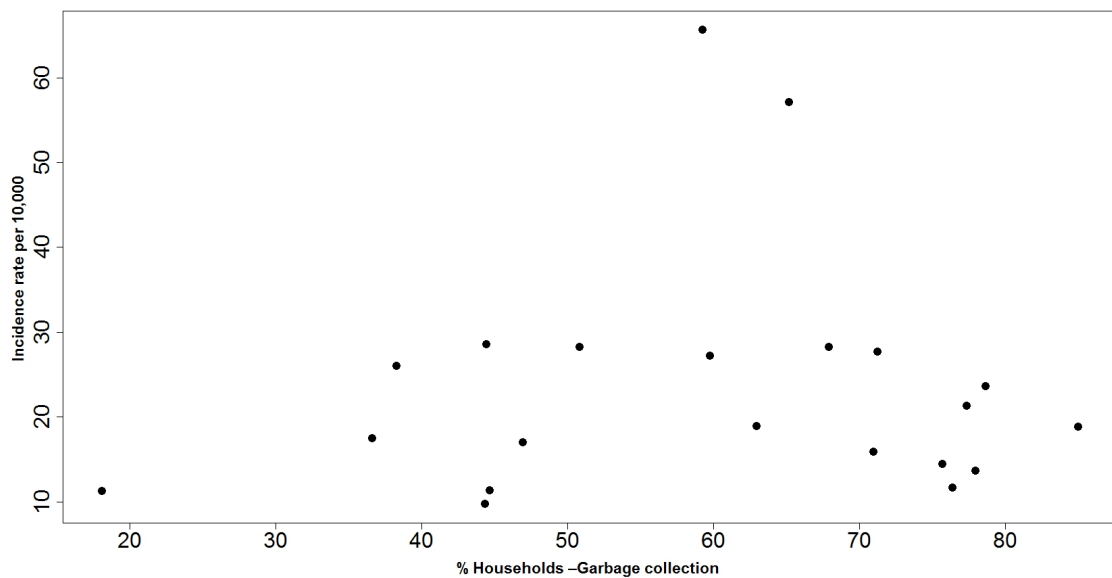


Figure A.1: Scatterplot of percentage of households which have a garbage collection service against incidence of dengue fever

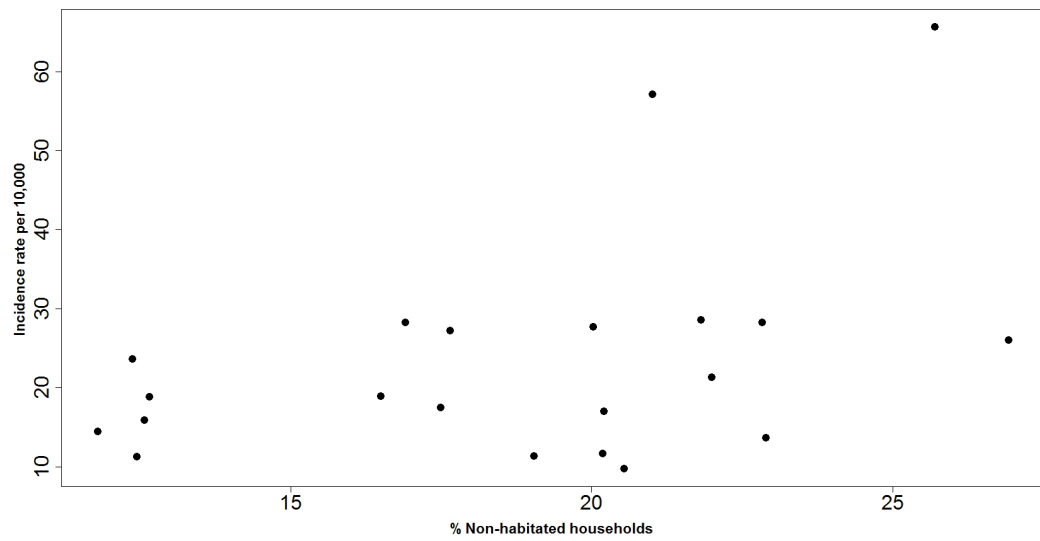


Figure A.2: Scatterplot of households which are unoccupied against incidence rate of dengue

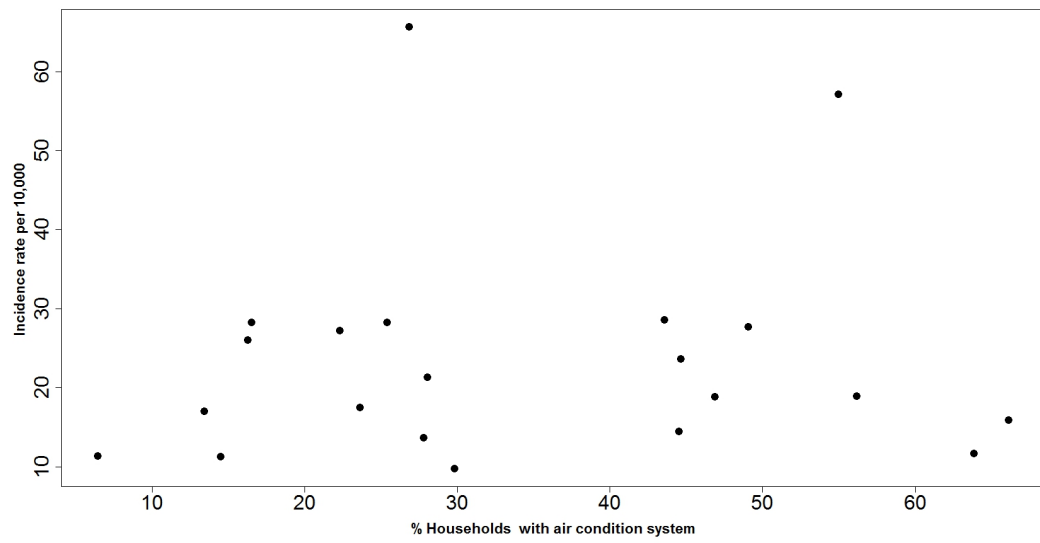


Figure A.3: Scatterplot of percentage of households with residential air conditioning equipment against incidence rate of dengue

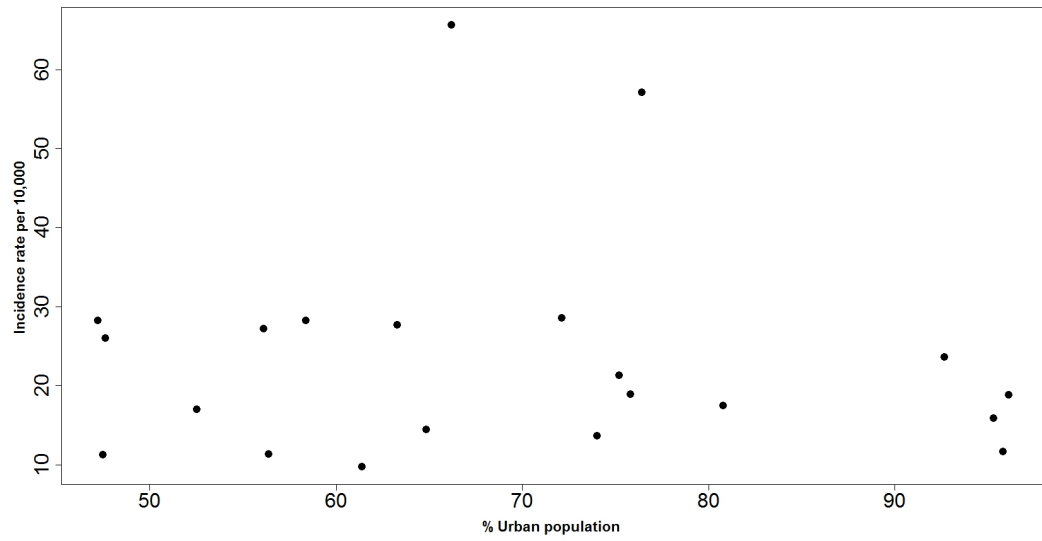


Figure A.4: Scatterplot of percentage of urban population against incidence of dengue fever

A.2 Multipanel of socioeconomic factors aggregated at municipality level

This approach shows potential associations amongst socio economic factors that might lead to issues of multicollinearity.

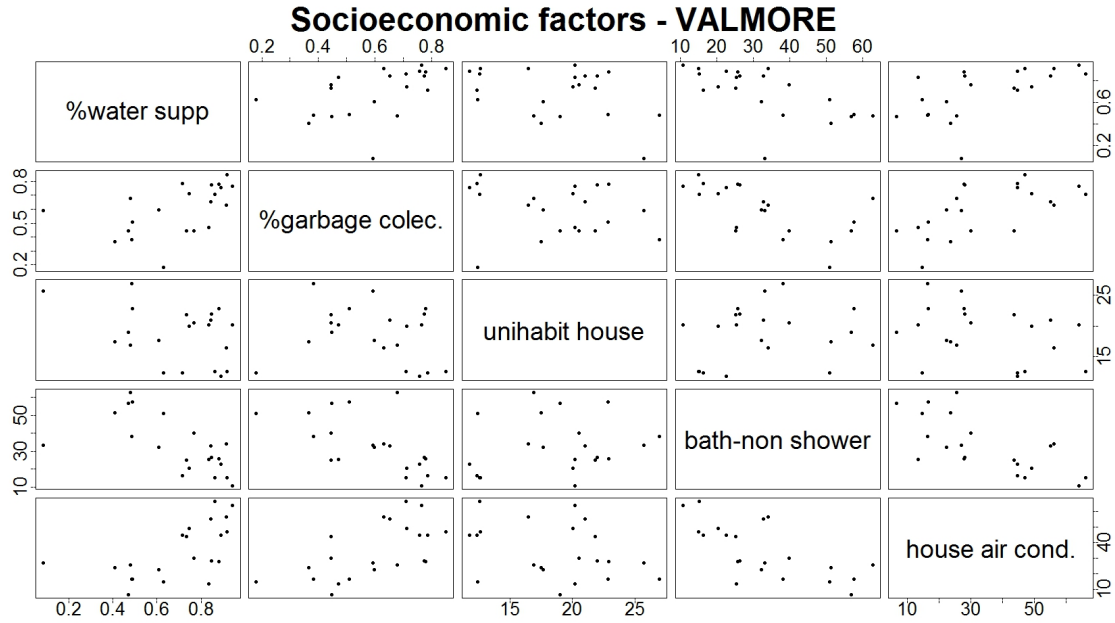


Figure A.5: Multipanel of socioeconomic variables in Valmore: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment

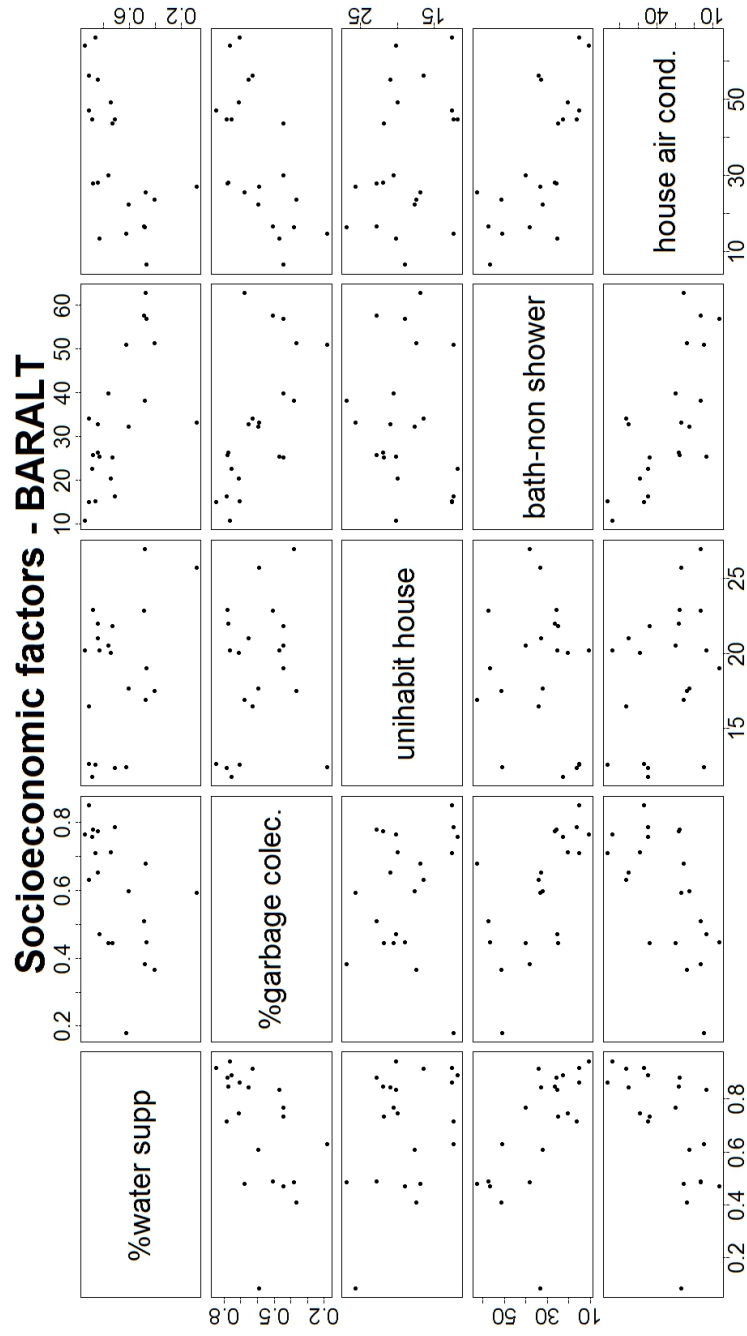


Figure A.6: Multipanel of socioeconomic variables in Baralt: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment

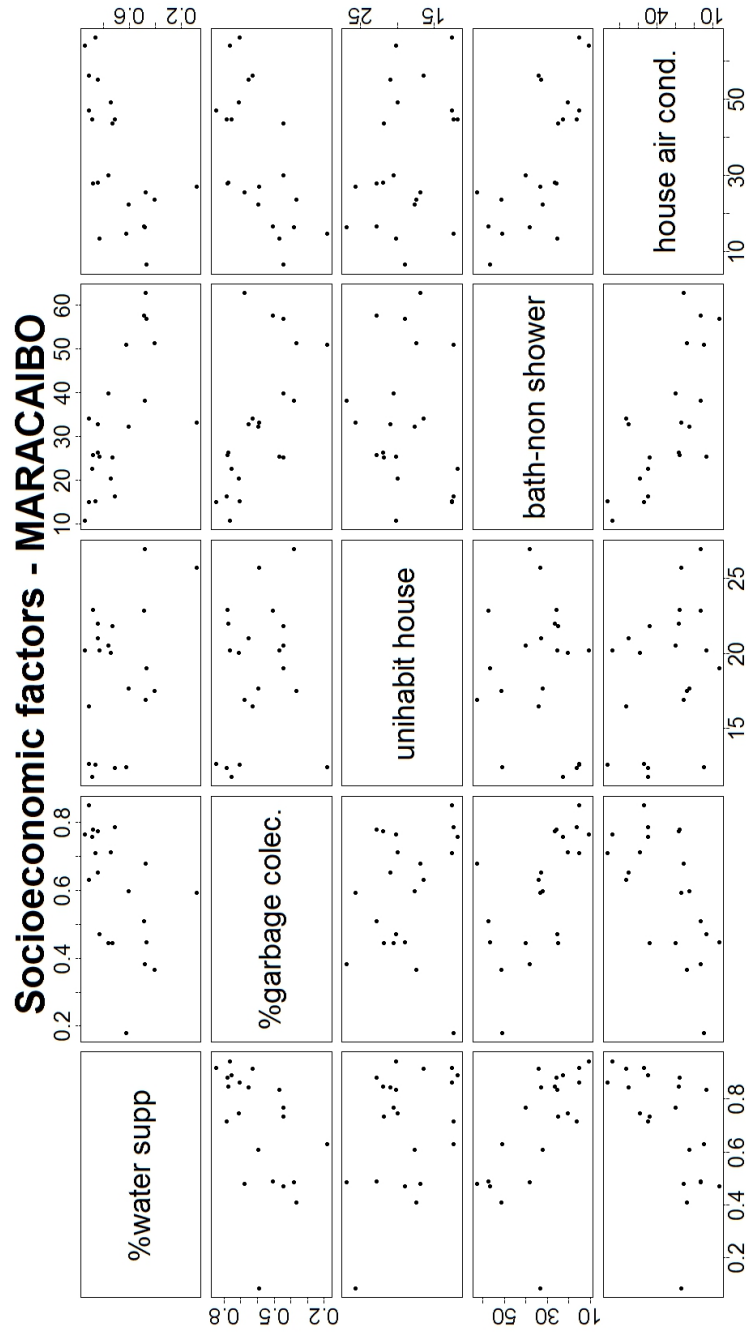


Figure A.7: Multipanel of socioeconomic variables in Maracaibo: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment

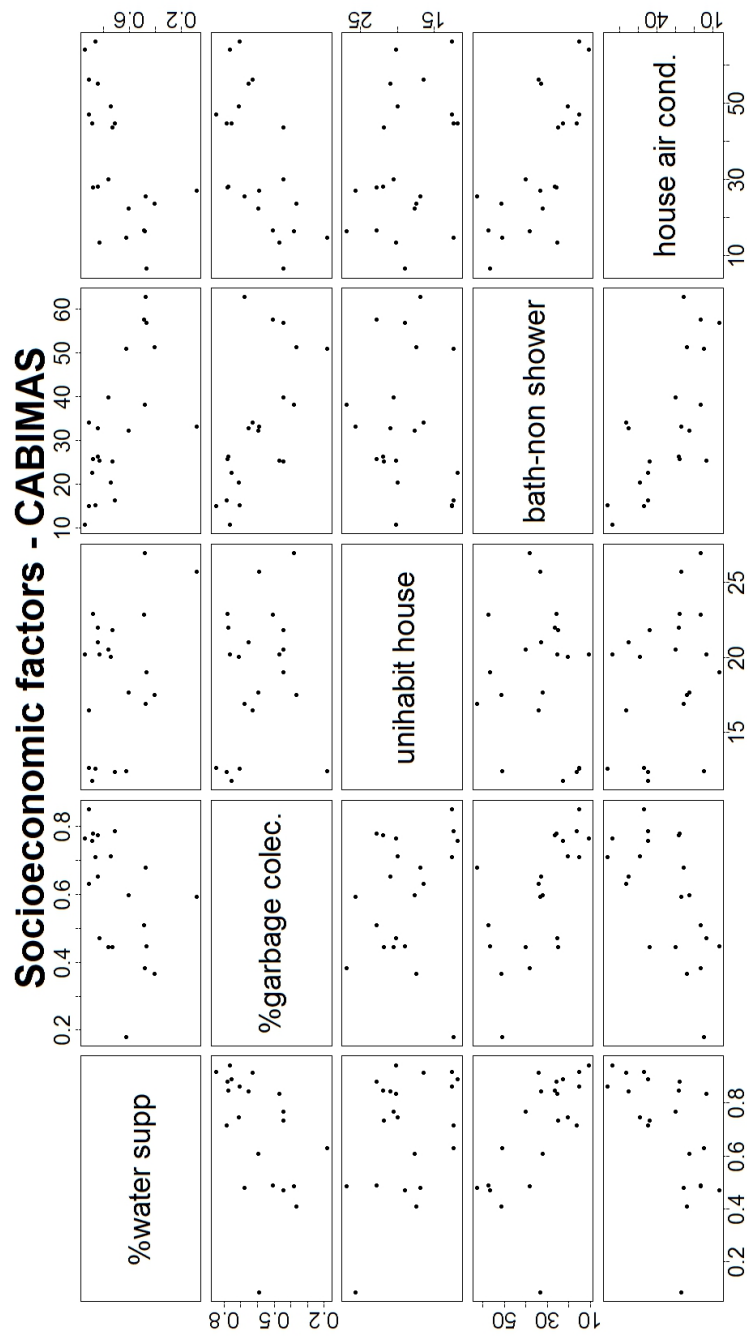


Figure A.8: Multipanel of socioeconomic variables in Cabimas: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment

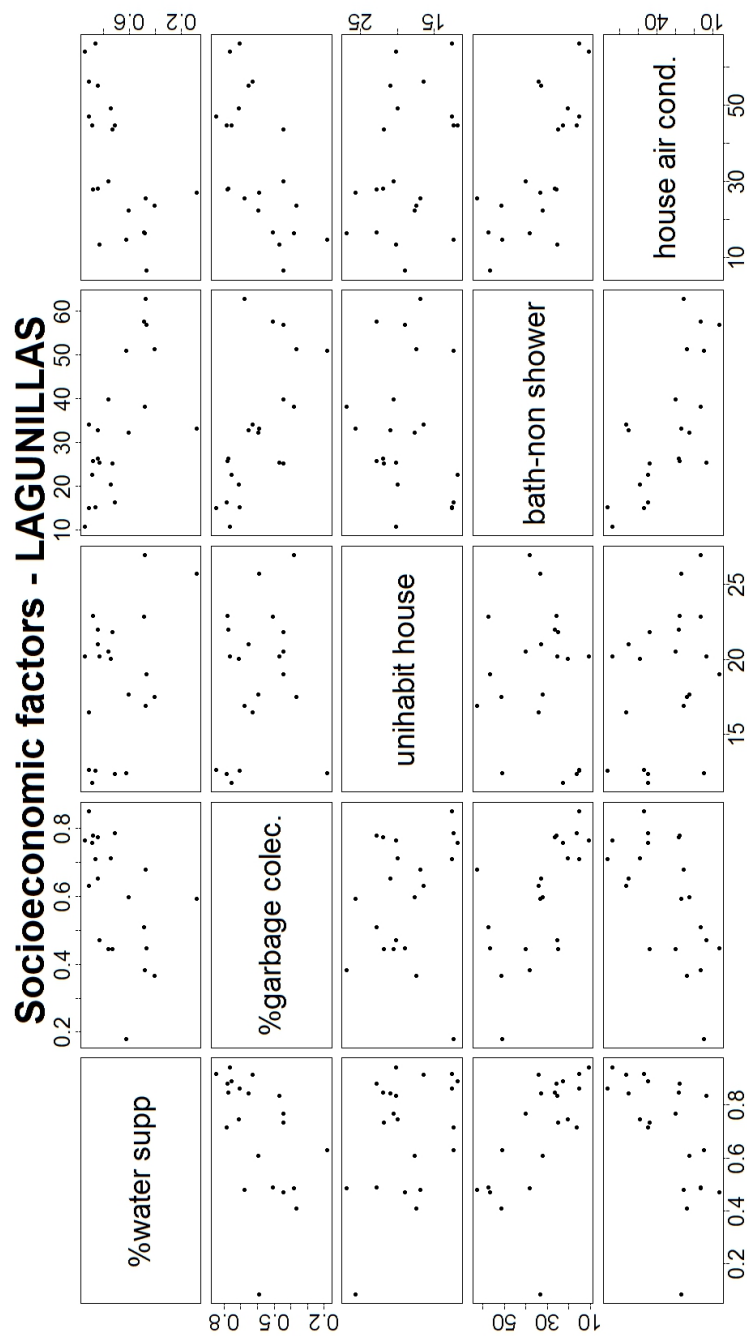


Figure A.9: Multipanel of socioeconomic variables in Lagunillas: (1) Percentage of households with connections to the water supply network, (2) Percentage of households which receive garbage collection service, (3) No. of households which are unoccupied, (4) No. of households including a bathroom without shower, (5) No. of households with residential air conditioning equipment

A.3 Multipanel scatterplots of socioeconomic factors in the whole of Zulia

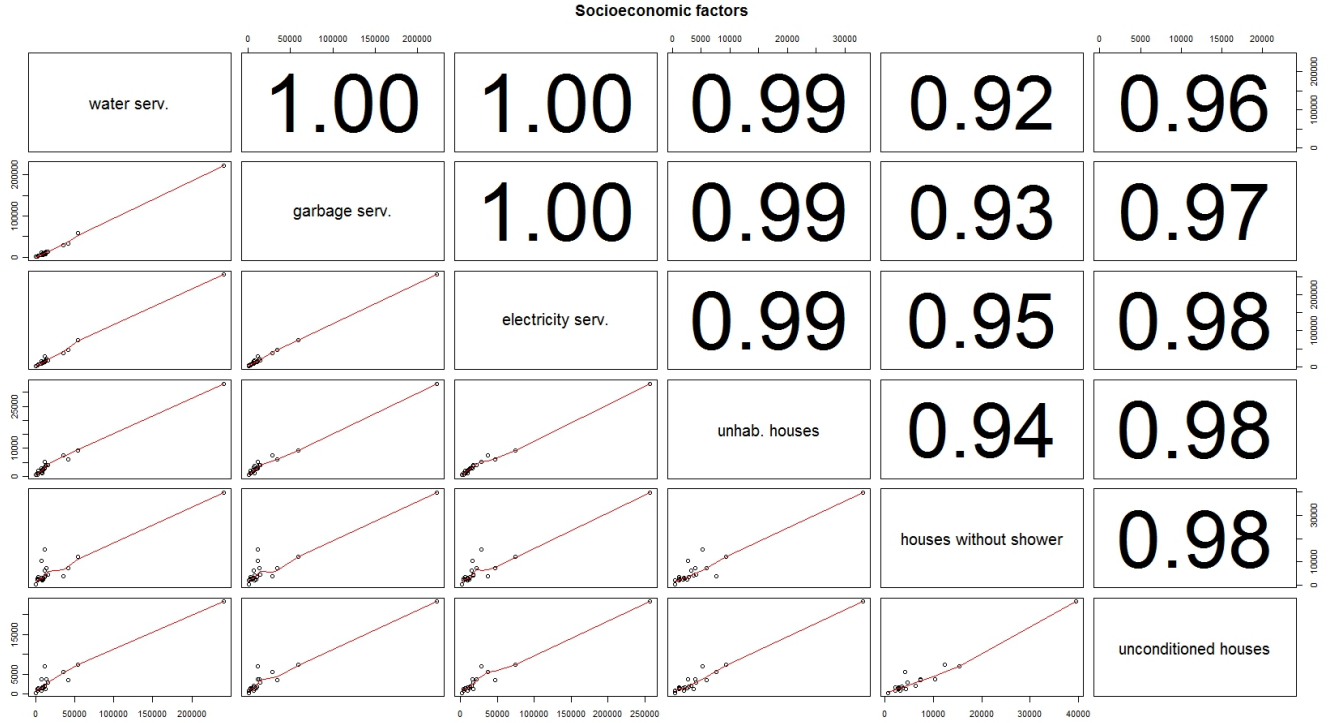


Figure A.10: Multipanel scatterplots of socioeconomic factors with Pearson's coefficients amongst them: (1) No. of households with access to the water supply network, (2) No. of homes which have garbage collection service, (3) No. of households with access to electricity, (4) No. of households which are unoccupied, (5) No. of households including a bathroom without basic shower, (6) No. of households without basic living standards

A.4 Scatterplot showing the association between temperature indexes

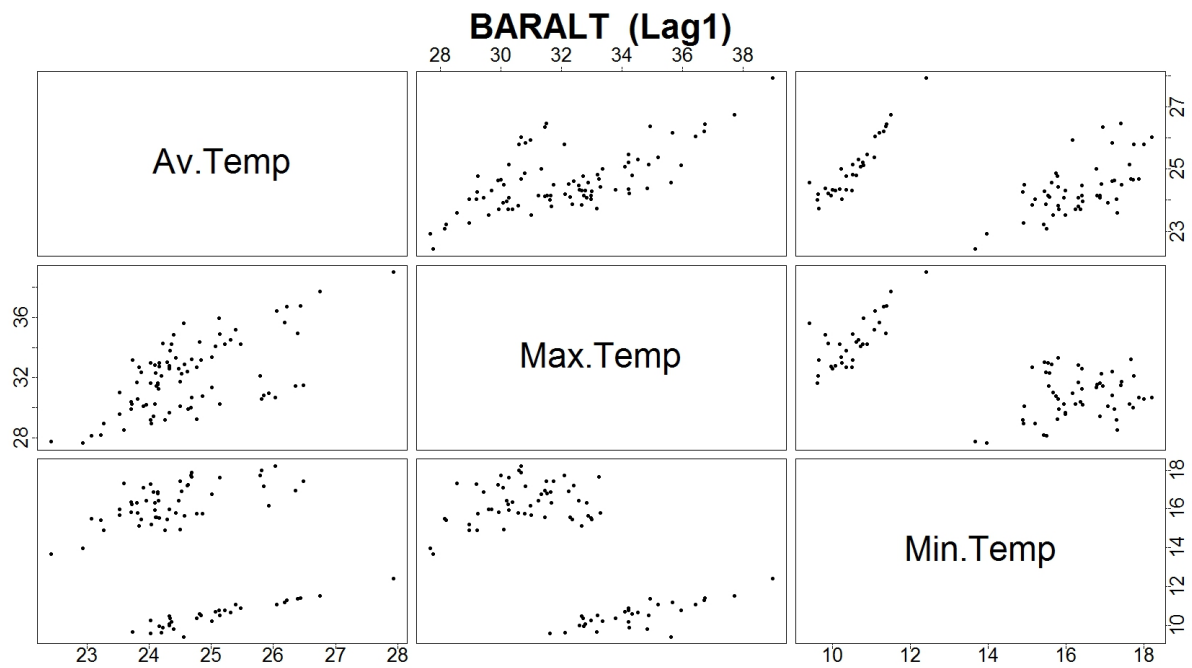


Figure A.11: Multipanel plots of temperature indexes with 1 month time lag in Baralt

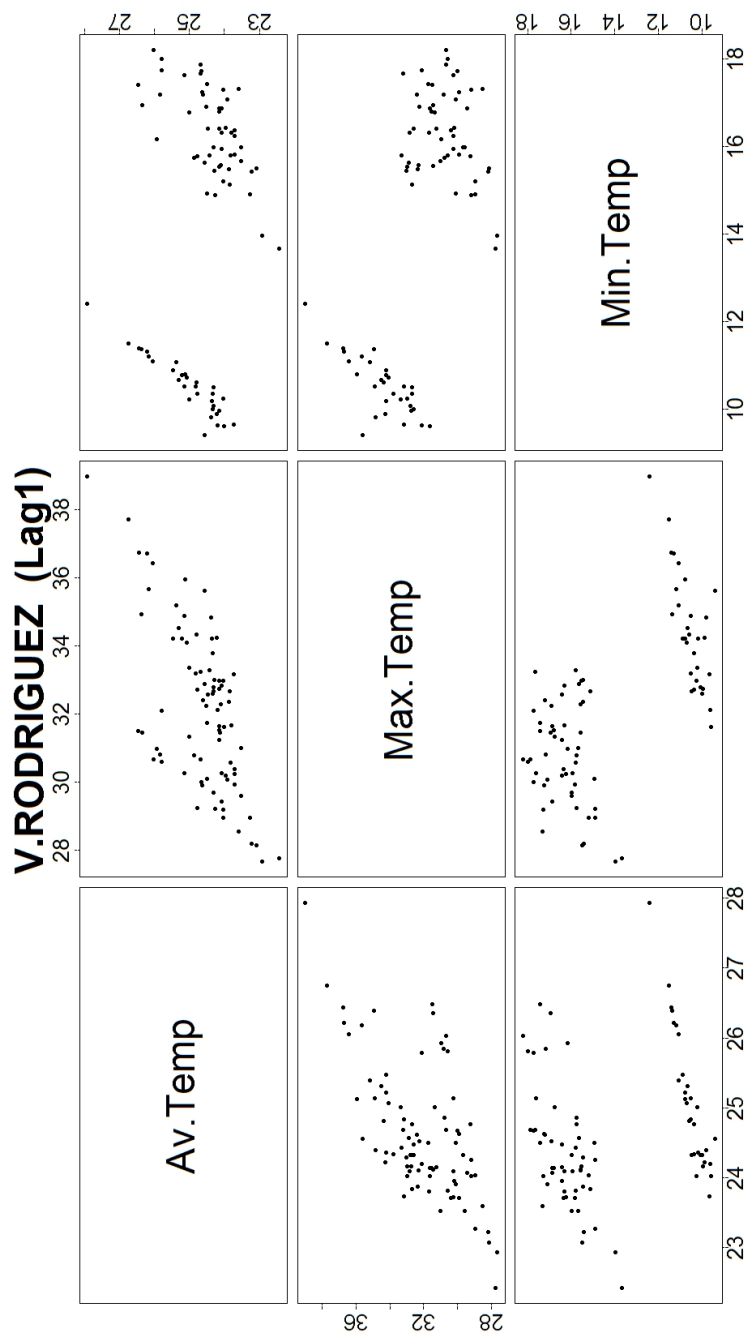


Figure A.12: Multipanel plots of temperature indexes with 1 month time lag in Valmore

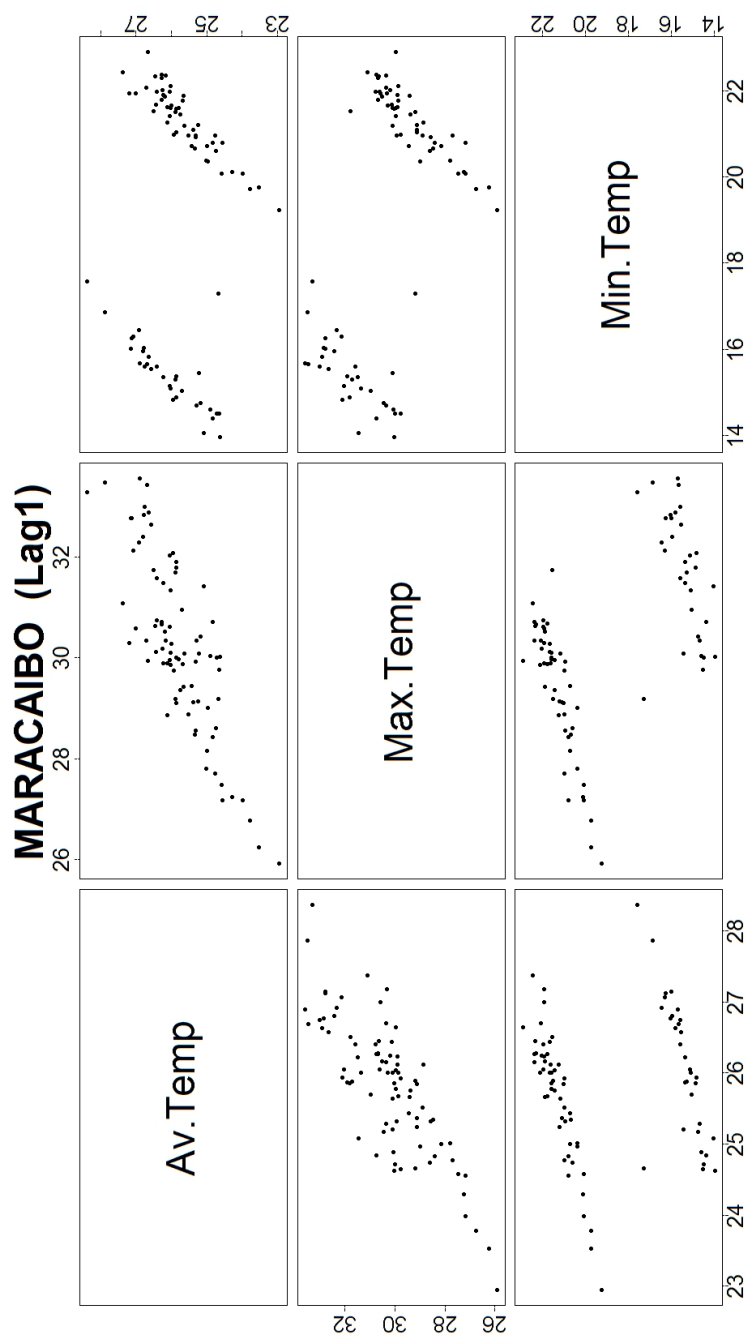


Figure A.13: Multipanel plots of temperature indexes with 1 month time lag in Maracaibo

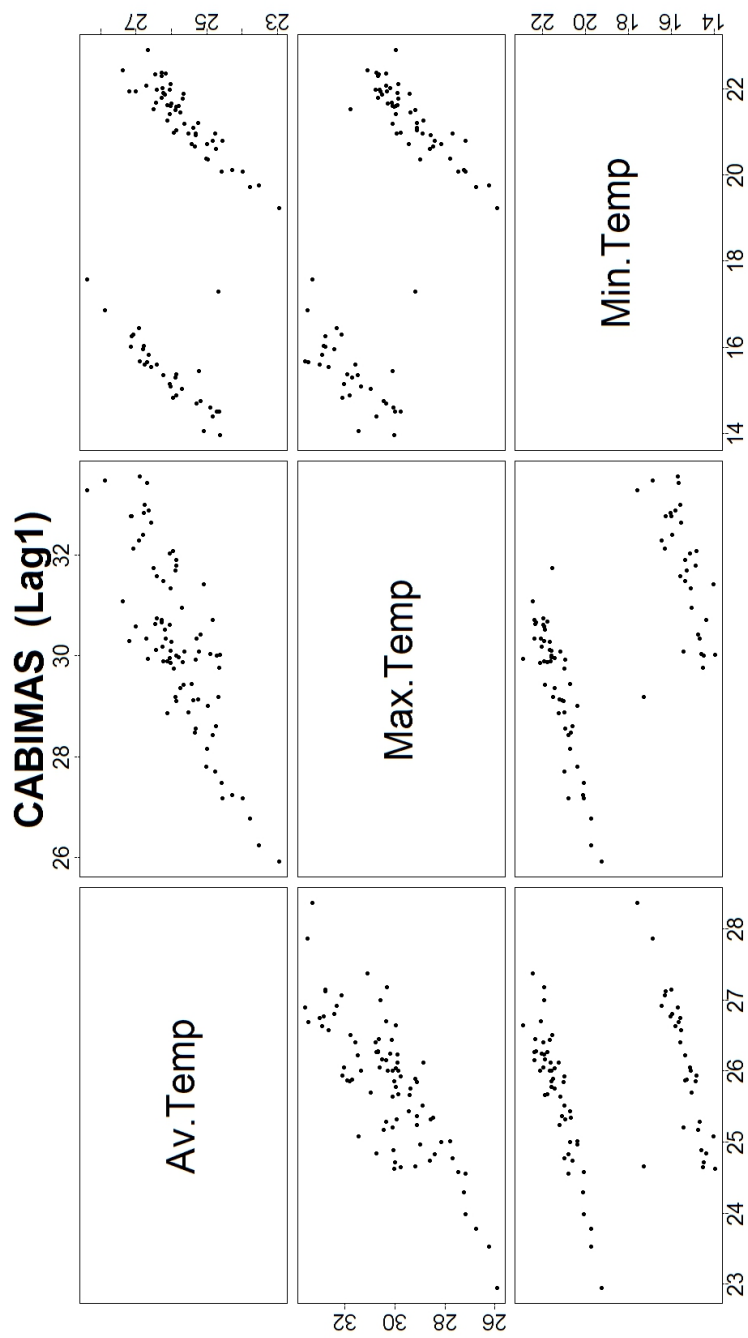


Figure A.14: Multipanel plots of temperature indexes with 1 month time lag in Cabimas

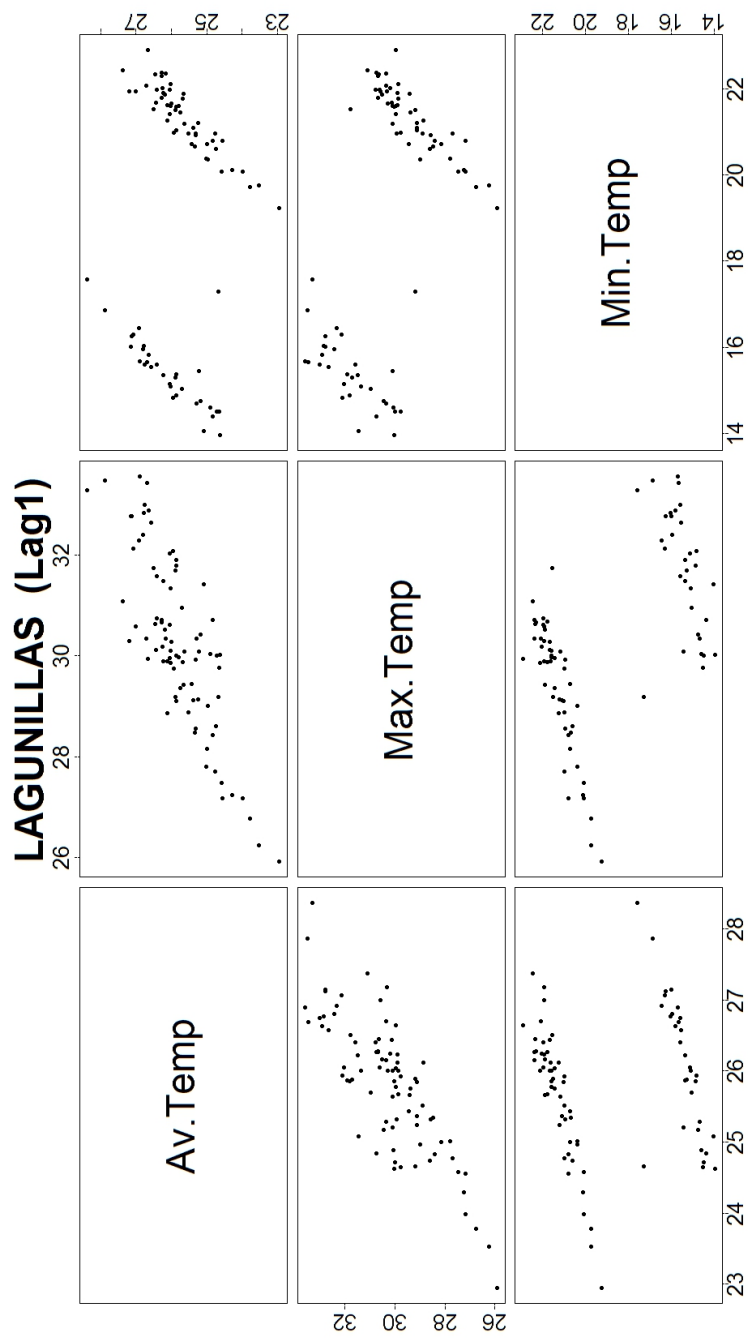


Figure A.15: Multipanel plots of temperature indexes with 1 month time lag in Lagunillas

Appendix B

Model specifications

B.1 Model specification via a frequentist approach

Model specification

The R-code using the frequentist approach for count data of dengue cases in the whole of Zulia state.

```
#####  
### Incidence rate per year & group of age & gender  
#####  
  
### Negative Binomial GLM: predictors as categorical variables  
  
newdat$year<- as.factor(newdat$year)  
newdat$agegroup<- as.factor(newdat$agegroup)  
newdat$sex<- as.factor(newdat$sex)  
  
glm(formula= count ~ year + agegroup + sex,  
family= neg binomial,  
data= newdat,  
offset = log(popsiZe))
```

B.2 Model specification via a Hierarchical Bayesian approach

First structure: Bayesian Hierarchical Poisson model

```
#####  
### Incidence rate per year in each agegroup  
#####  
  
model  
{  
  
  for (i in 1:70){ ##### number of observations  
    log(mu[i])<-log(popsize[i])+intercept+  
      yeargrp2[i]*beta[1]+  
      yeargrp3[i]*beta[2]+  
      yeargrp4[i]*beta[3]+  
      yeargrp5[i]*beta[4]+  
      yeargrp6[i]*beta[5]+  
      yeargrp7[i]*beta[6]+  
  
      agegrp2[i]*beta[7]+  
      agegrp3[i]*beta[8]+  
      agegrp4[i]*beta[9]+  
      agegrp5[i]*beta[10]+eps[i]  
  
    count[i]~dpois(mu[i]) ##### response modelb  
    eps[i]~dnorm(0.0,eps.tau)  
  
  }  
  
#####  
##### Priors distributions  
#####  
  
  eps.tau~dgamma(0.001,0.001) ##### Vague prior information  
  intercept~dnorm(0,1.0E-6) # Flat prior for the intercept
```

```

beta[1]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[2]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[3]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[4]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[5]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[6]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[7]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[8]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[9]~dnorm(0,1.0E-6)      #####   Vague normal priors for beta
parameters
beta[10]~dnorm(0,1.0E-6)     #####   Vague normal priors for beta
parameters

}

```

Second approach: Bayesian Hierarchical Negative Binomial model

```

#####
# Incidence rate per year in each agegroup #
#####

model

{

for (i in 1:70){    # amount of data

count[i]~dpois(mustar[i])  # response variable as product of two

```

```

parameters
mustar[i]<-rho[i]*mu[i]      # usual single Poisson parameter as a product
log link
log(mu[i])<-log(popsize[i])+intercept+

  yeargrp2[i]*beta[1]+
  yeargrp3[i]*beta[2]+
  yeargrp4[i]*beta[3]+
  yeargrp5[i]*beta[4]+
  yeargrp6[i]*beta[5]+
  yeargrp7[i]*beta[6]+
  agegrp2[i]*beta[7]+
  agegrp3[i]*beta[8]+
  agegrp4[i]*beta[9]+
  agegrp5[i]*beta[10]+

  rho[i]~dgamma(alpha,alpha)  # overdispersion parameter

}
# Priors distributions
intercept~dnorm(0,1.0E-6) # Flat prior for the intercept
beta[1]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[2]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[3]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[4]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[5]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[6]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[7]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[8]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[9]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters
beta[10]~dnorm(0,1.0E-6) # Vague normal priors for beta parameters

alpha<-exp(logalpha)
logalpha~dnorm(0,1.0E-6)

}

```

B.3 Analysis of results from the Bayesian approach

The visualization of the autocorrelation functions of $b[1]$, $b[2]$ and $b[7]$ parameters from the Hierarchical Bayesian Poisson version, via WinBUGS are displayed below.

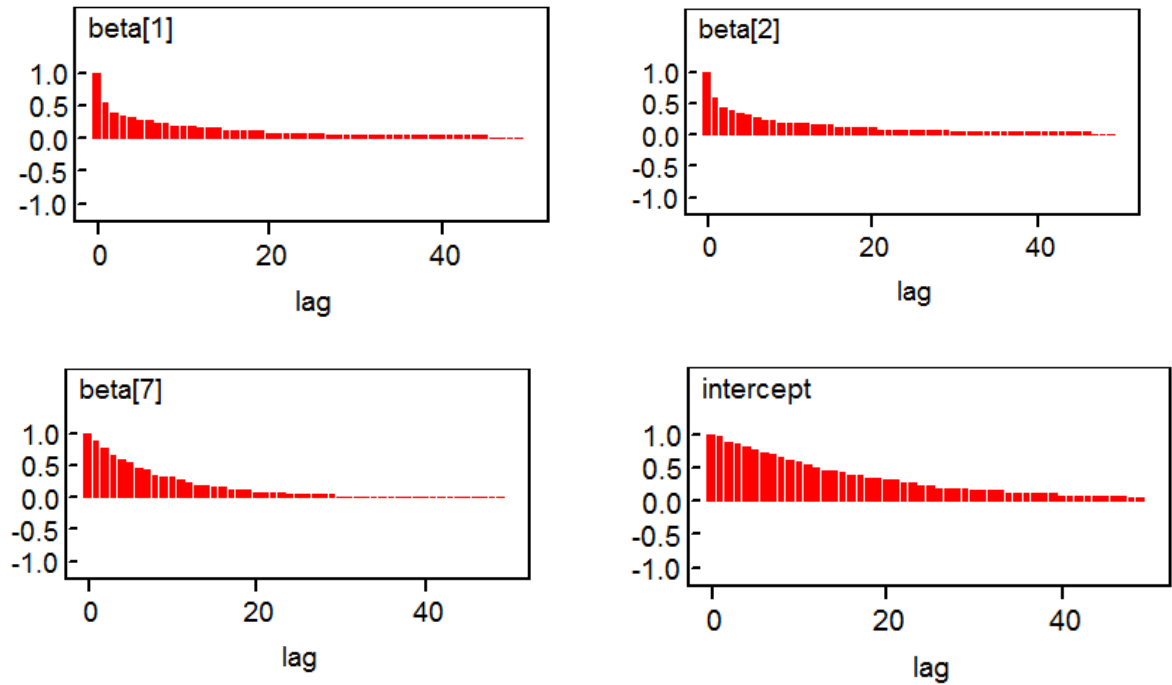


Figure B.1: WinBUGS outputs of dengue incidence in Zulia state. Autocorrelation functions

And finally the visual diagnostic revealed that the bivariate posterior scatter plot around the mean, was randomly distributed for all possible combination of parameters, meaning that the convergence was reached. To this end the patterns for parameters $\beta[1]$ against $\beta[7]$ and $\beta[3]$ against $\beta[7]$ were plotted.

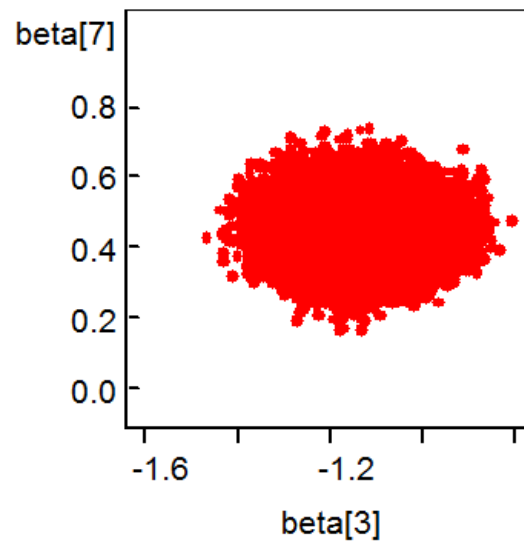
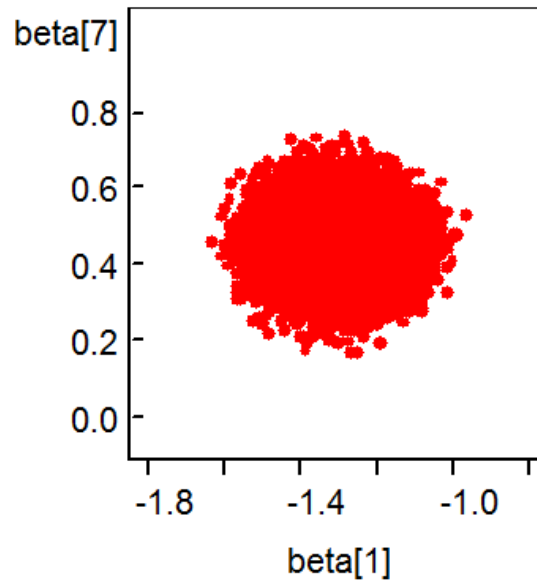


Figure B.2: Bivariate posterior scatter plots

Appendix C

GAM: Spatio-Temporal modelling

C.1 First stage: GAMs for lagged effects

Model specification

Three resulting models for the final selection process regarding the lagged effects of climatic covariates were specified as a Negative Binomial GAM structure with smooth terms for average rainfall and maximum temperature at 1 month delay respectively. In conjunction with the El Nino3.4 at 6 months delay. A set of climatic factors treated with smooth terms across the five regions which comprises Zulia state (used as a factor). The first model was performed with free restriction on the degree of freedom (df) coded in *mgcv* as `k=-1`. The second model was specified with smooth terms fixed at `k=5` degree of freedom. And the last model was settled with an amount of smoothing at `k=3`.

The R-codes from these models are listed below, using a data set aggregated from January 2002 to December 2008.

```
#####  
##### Climatic combination of factors lag time at 1 month  
##### MODEL: MAX TEMPERATUE (lag1)+AV.RAINFALL (lag1)
```



```
#####
#### Free restriction on degree of freedom

newdat$region<-as.factor(newdat$region)

gam(obsdengue~ region+s(maxtempclag1,fx=FALSE,k=-1,bs="cr")+
      s(meanrainfratelag1mm, fx=FALSE,k=-1,bs="cr")+
      s(nino3.4lag6, fx=FALSE,k=-1,bs="cr")+
      +
      s(altitude, fx=FALSE,k=-1,bs="cr")+
      s(perhwater, fx=FALSE,k=-1,bs="cr")+
      s(area, fx=FALSE,k=-1,bs="cr"),
      method="GCV.Cp",offset=log(popsize),
      family=negbin(1),data=newdat)
```

```
#####
#### Specification of Model with k=5
# MODEL: MAX TEMPERATUE (lag1)+AV.RAINFALL (lag1)
#####
# Fixed degree of freedom at 5

library(mgcv)
library(MASS)
newdat$region<-as.factor(newdat$region)

gam(obsdengue~ region+
      s(maxtempclag1,fx=FALSE,k=5,bs="cr")+
      s(meanrainfratelag1mm, fx=FALSE,k=5,bs="cr")+
      s(nino3.4lag6, fx=FALSE,k=5,bs="cr")+
      s(altitude, fx=FALSE,k=5,bs="cr")+
      s(perhwater, fx=FALSE,k=5,bs="cr")+
      s(area, fx=FALSE,k=5,bs="cr"),
      method="GCV.Cp",offset=log(popsize),
      family=negbin(1),data=newdat)
```

```
#####
```

```
##### Specification of Model with k=3
# MODEL: MAX TEMPERATUE (lag1)+AV.RAINFALL (lag1)
#####
# Fixed degree of freedom at 3

gam(obsdengue~ region+
      s(maxtempclag1, fx=FALSE, k=3, bs="cr")+
      s(meanrainfratelag1mm, fx=FALSE, k=3, bs="cr")+
      s(nino3.4lag6, fx=FALSE, k=3, bs="cr")+
      s(altitude, fx=FALSE, k=3, bs="cr")+
      s(perhwater, fx=FALSE, k=3, bs="cr")+
      s(area, fx=FALSE, k=3, bs="cr"),
      method="GCV.Cp", offset=log(popsizer),
      family=negbin(1), data=newdat)
```

Model Validation

The Pearson and Deviance residuals were plotted vs. the fitted values from the optimal GAM structure with free restriction on the degree of freedom, using a data set from the whole of Zulia.

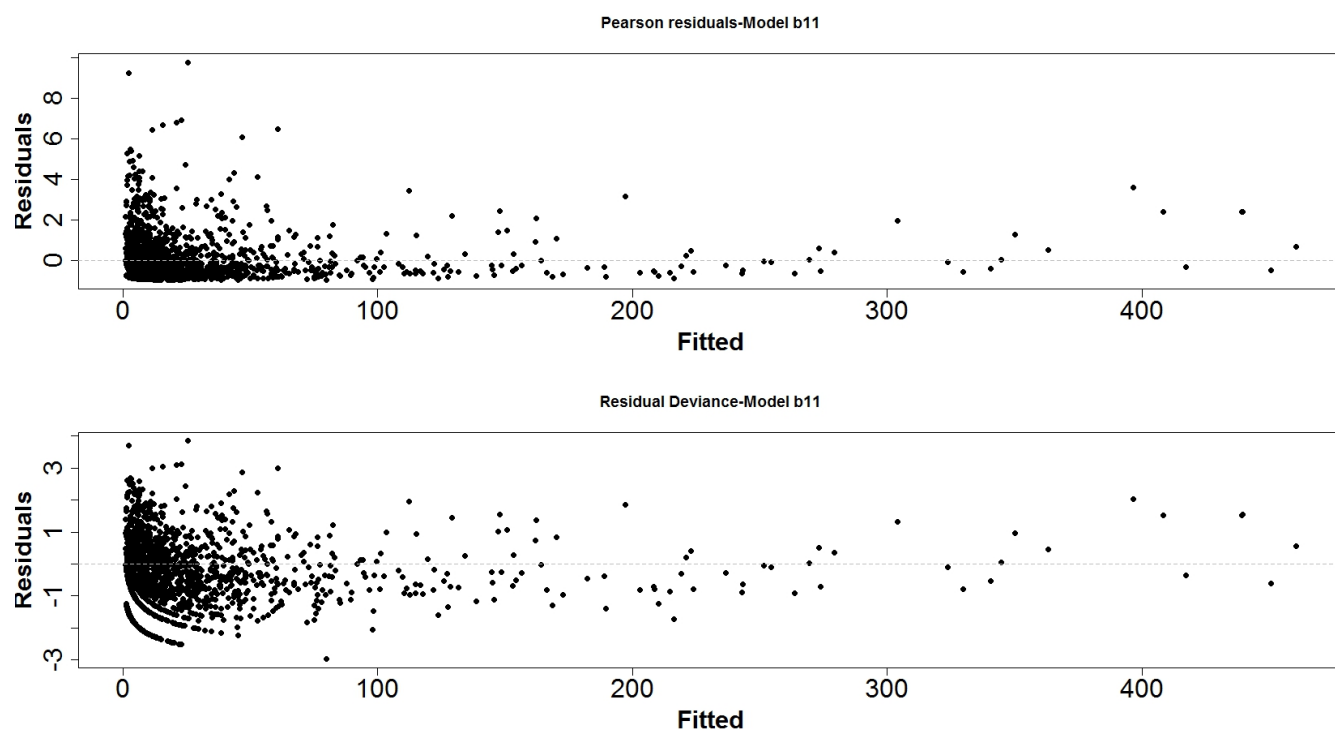


Figure C.1: Pearson and Deviance residual against fitted values from the optimal model with free restriction on the degree of freedom

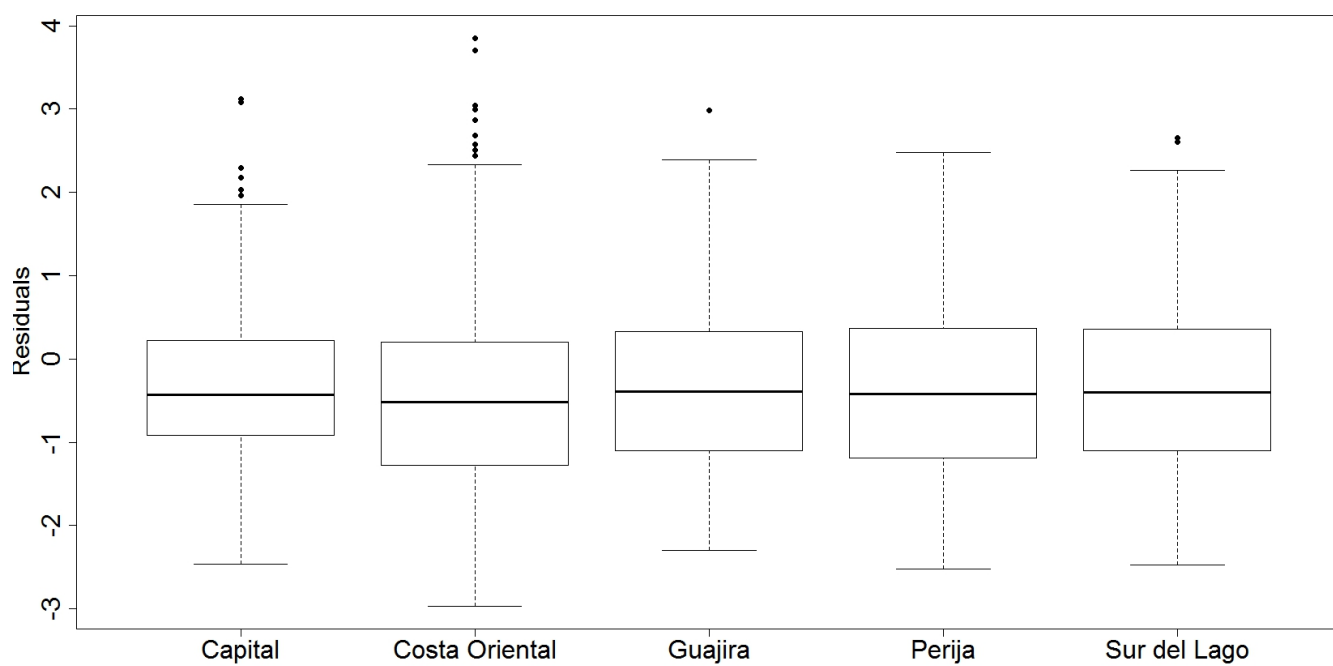


Figure C.2: Deviance residual split up at regional level

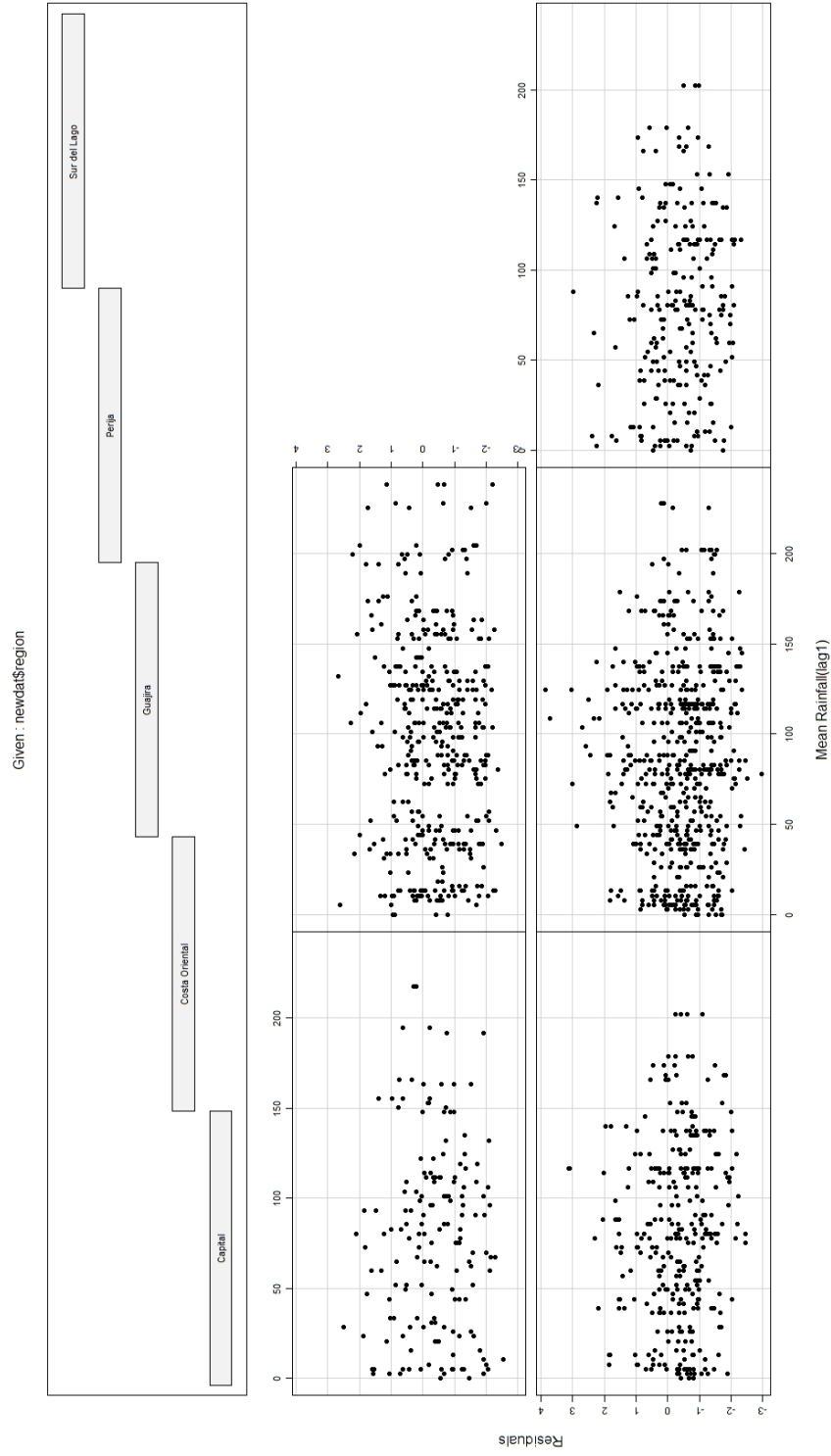


Figure C.3: Deviance residual against mean rainfall from the optimal model with free restriction on the degree of freedom

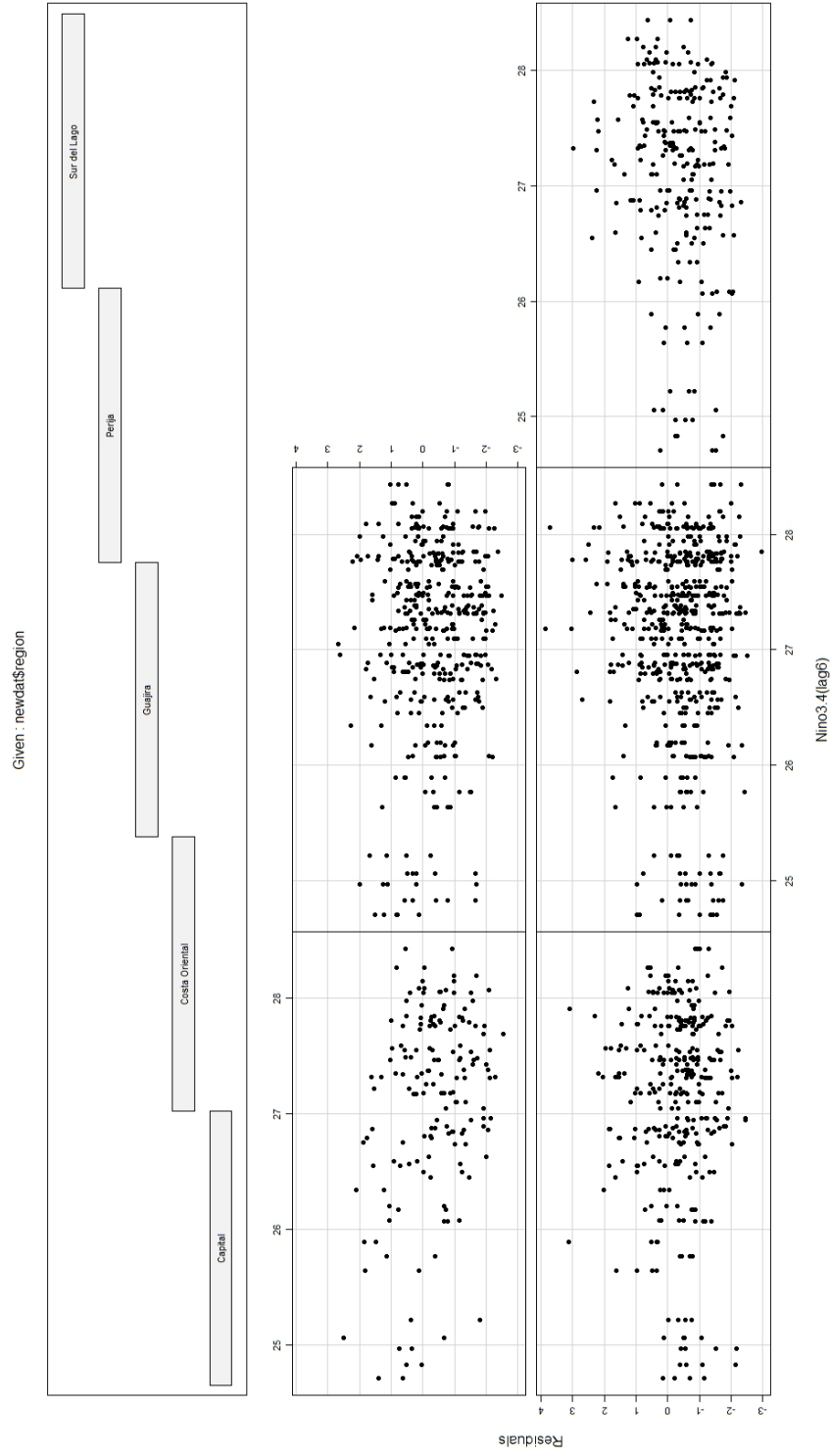


Figure C.4: Deviance residual against the El Nino3.4 from the optimal model with free restriction on the degree of freedom

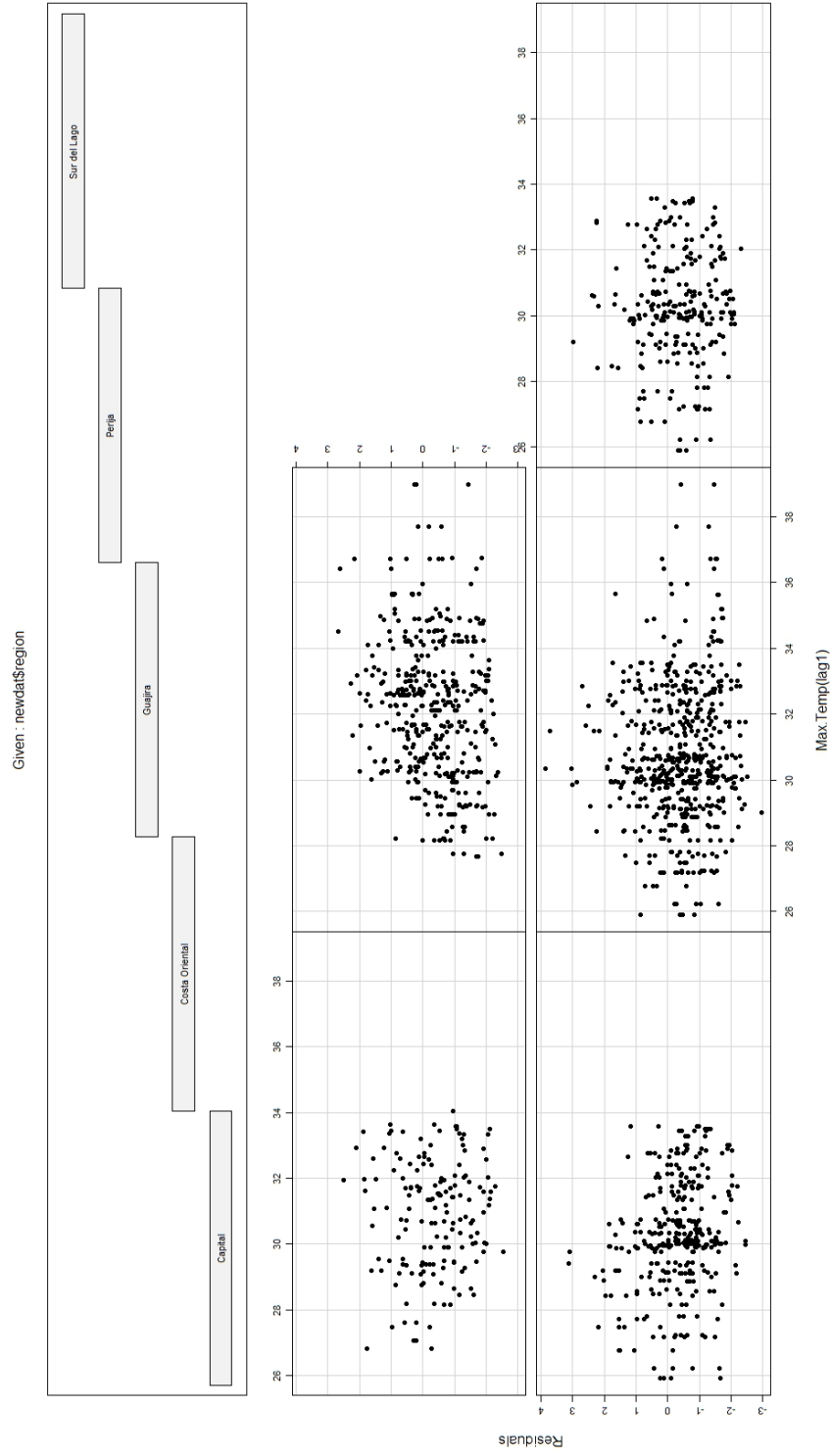


Figure C.5: Deviance residual against maximum temperature from the optimal model with free restriction on the degree of freedom

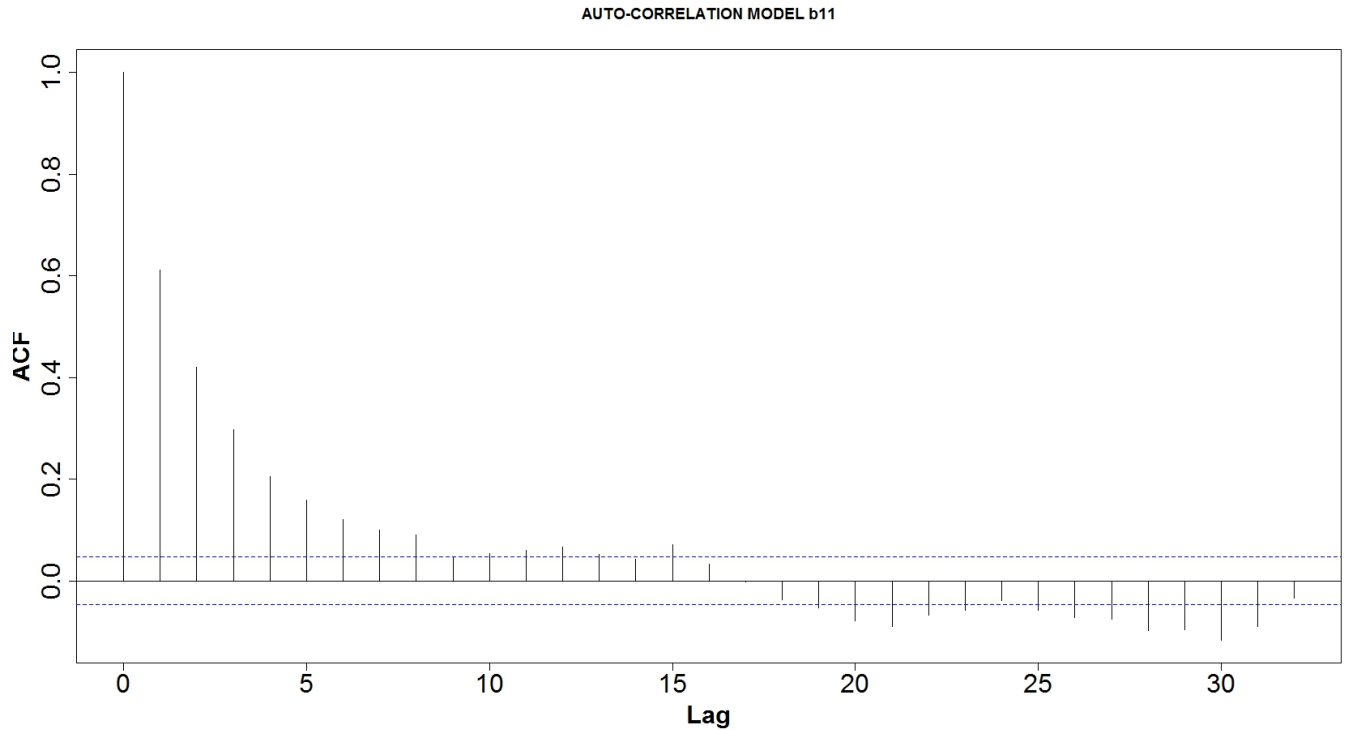


Figure C.6: ACF from optimal model with free restriction on the degree of freedom

C.2 Second stage: GAMM with municipality as a random effect

Model specification

A new approach was performed owing to some limitations of the previous one. The R-code using municipality as a random effect with smooth terms for seasonality and long term trend were included. Also, the non-climatic factors were defined in a categorical form.

```
#####
##### GAMM model with random effect #####
#####
newdat$municipality<-as.factor(newdat$municipality)
```



```

newdat$altgrp <-as.factor(newdat$altgrp)
newdat$perhwatergrp <-as.factor(newdat$perhwatergrp)

gam(obsdengue~s(municipality,bs="re")+altgrp+perhwatergrp+
      s(timeserie,bs="cs")+s(month,bs="cc
      ")+
      s(maxtempclag1,bs="cs",k=5)+
      s(meanrainfratelag1mm,bs="cs",k=5)+
      s(nino3.4lag6,bs="cs",k=5),
      offset=log(popsize), family=poisson,
      data=newdat,scale=-1,gamma=1.4)

```

Validation of the GAMM using random effect at municipal level

The following plots show the fitted values vs the Deviance residuals along the 21 municipalities which comprises Zulia state: Maracaibo, Cabimas, Lagunillas, Colon, Catatumbo, Urdaneta, Francisco, Mara, Miranda, Rosario, Baralt, Sucre, Pulgar, Machiques, Bolivar, Padilla, Lossada, Semprum, Rita, Rodriguez and Paez.

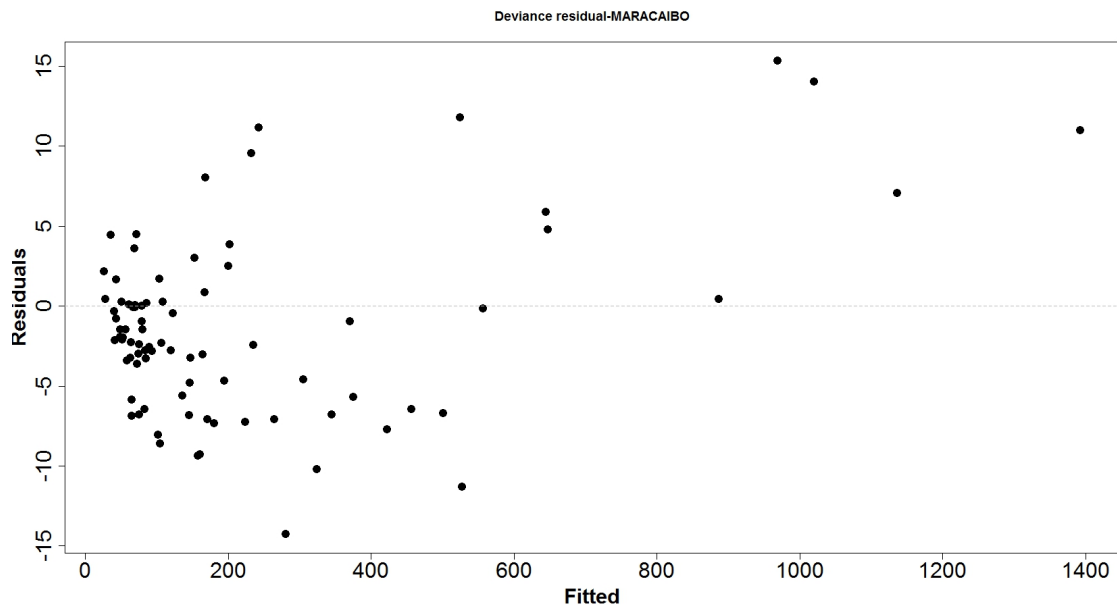


Figure C.7: Deviance of GAMM with random effect - Maracaibo

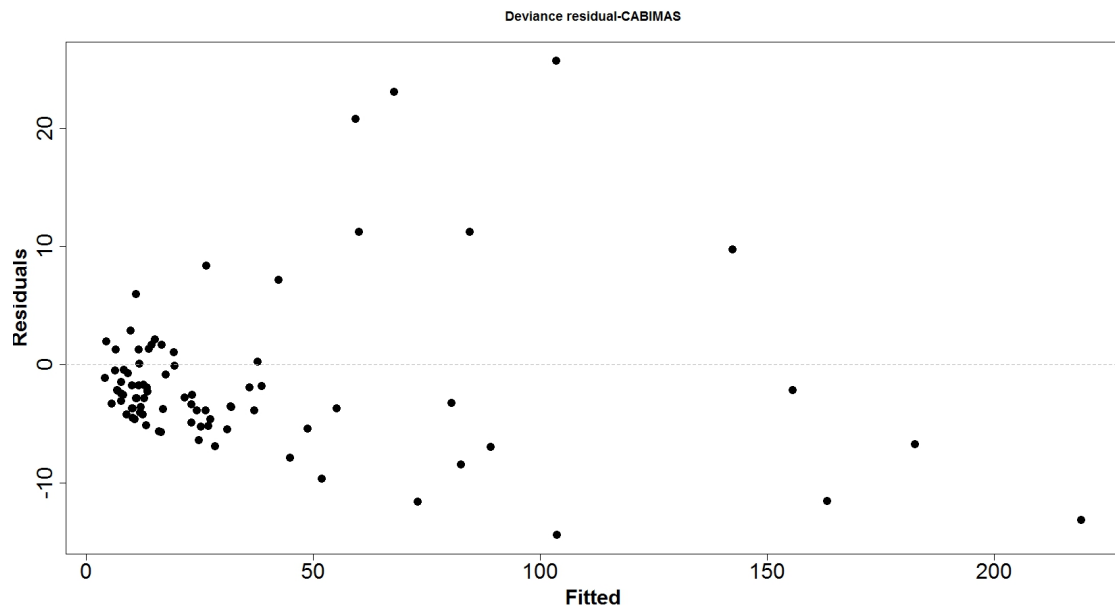


Figure C.8: Deviance of GAMM with random effect - Cabimas

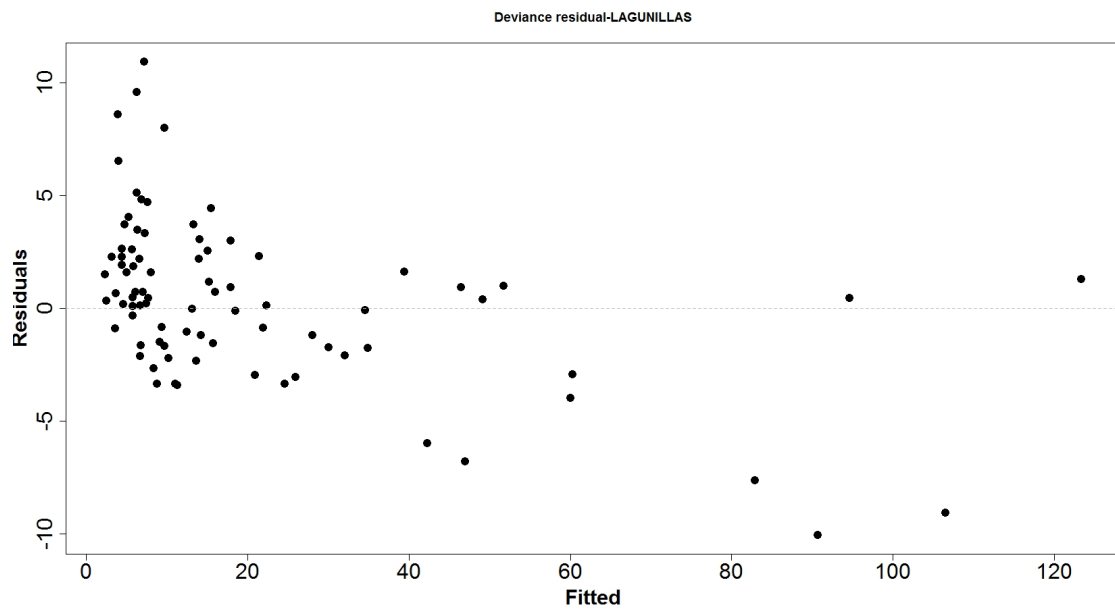


Figure C.9: Deviance of GAMM with random effect - Lagunillas

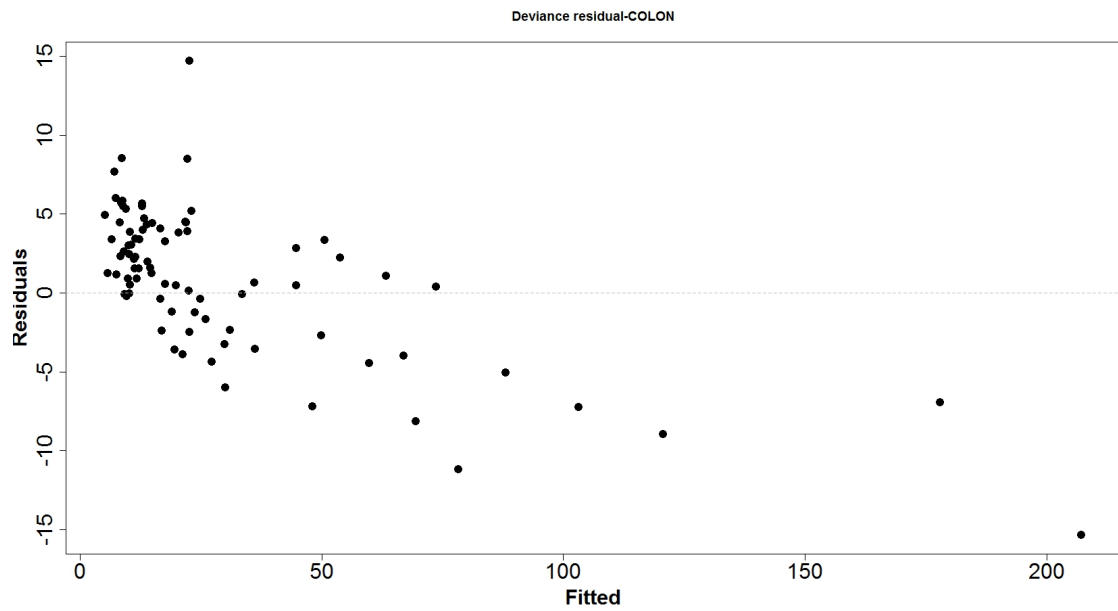


Figure C.10: Deviance of GAMM with random effect - Colon

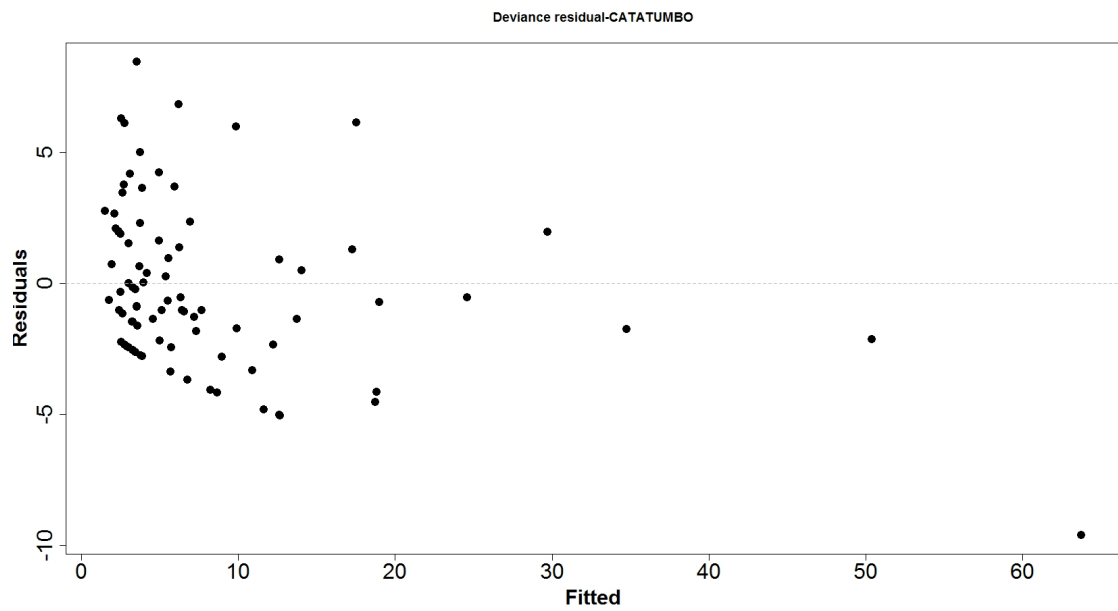


Figure C.11: Deviance of GAMM with random effect - Catatumbo

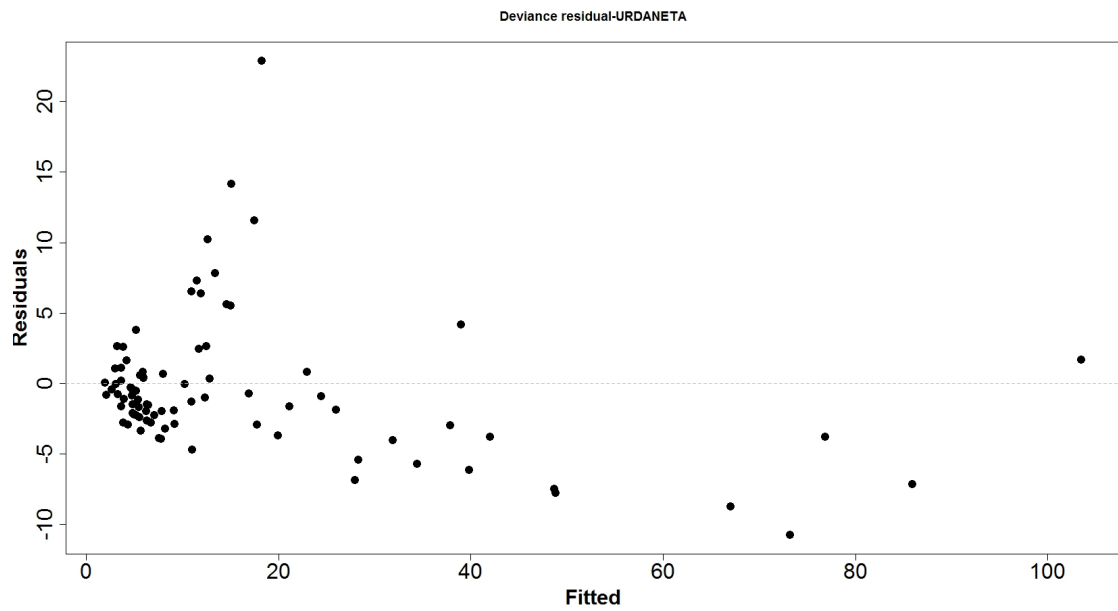


Figure C.12: Deviance of GAMM with random effect - Urdaneta

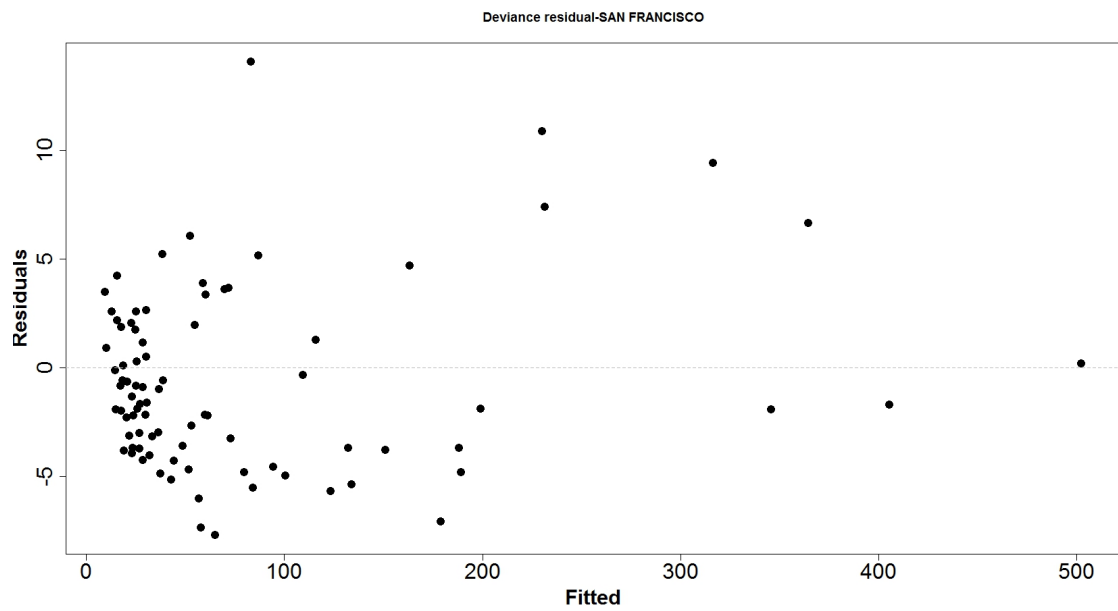


Figure C.13: Deviance of GAMM with random effect - Francisco

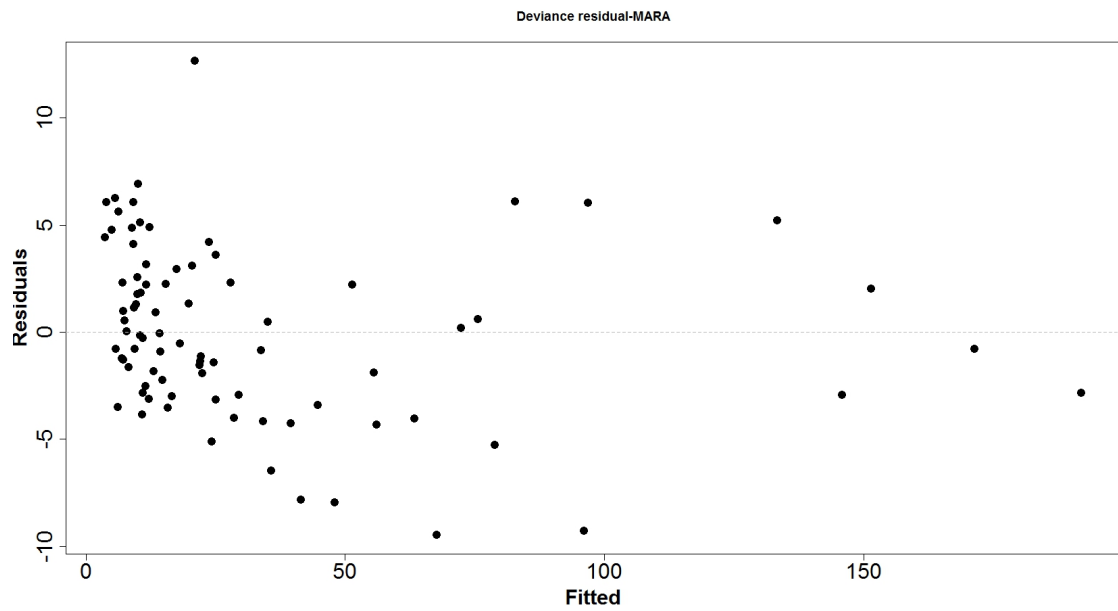


Figure C.14: Deviance of GAMM with random effect - Mara

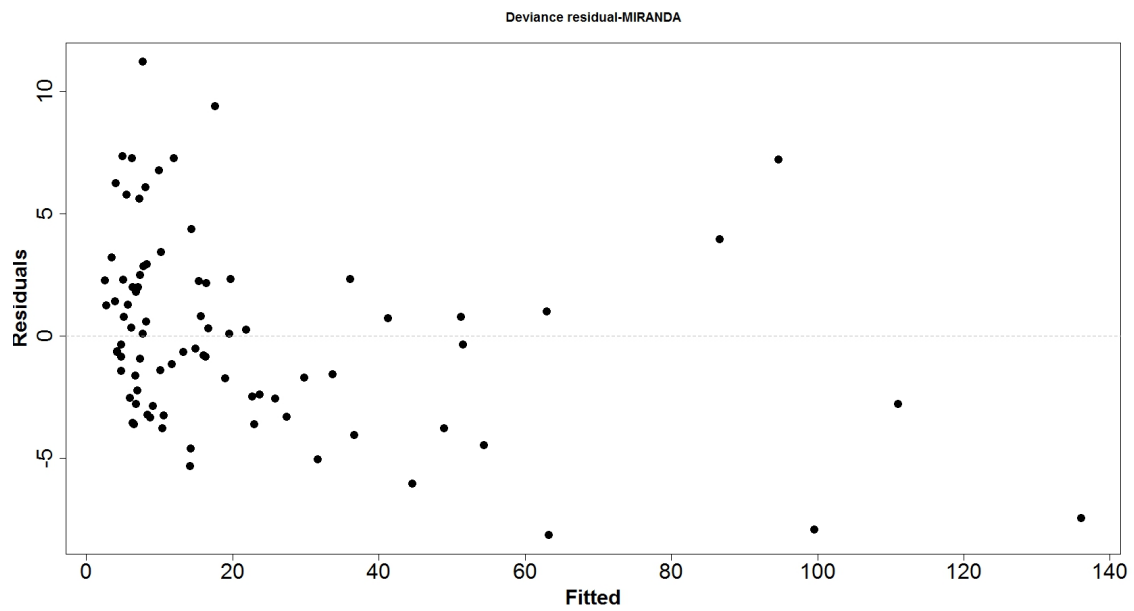


Figure C.15: Deviance of GAMM with random effect - Miranda

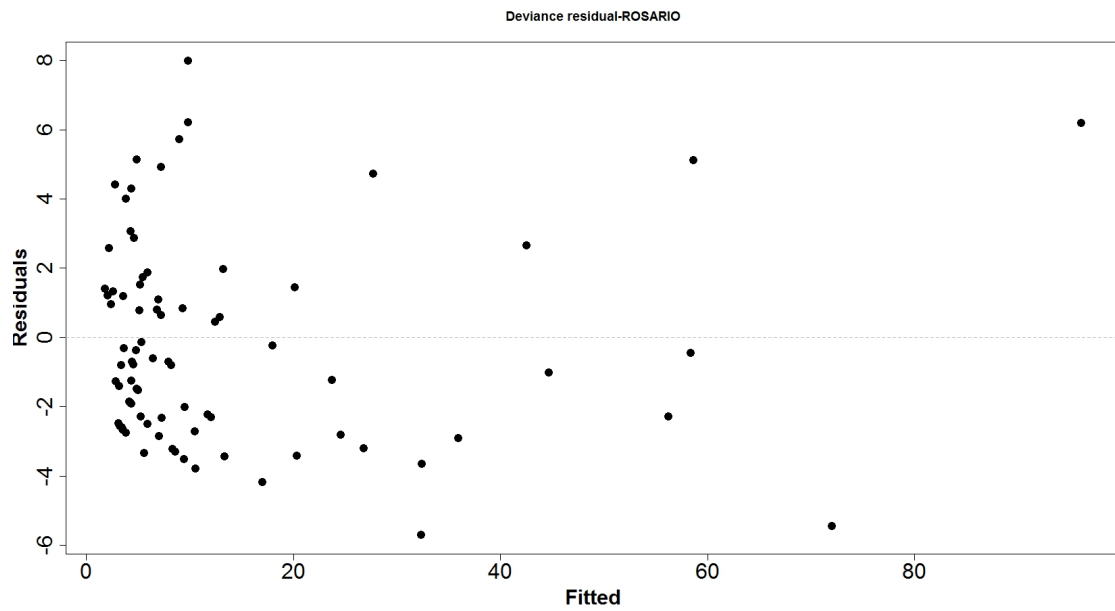


Figure C.16: Deviance of GAMM with random effect - Rosario

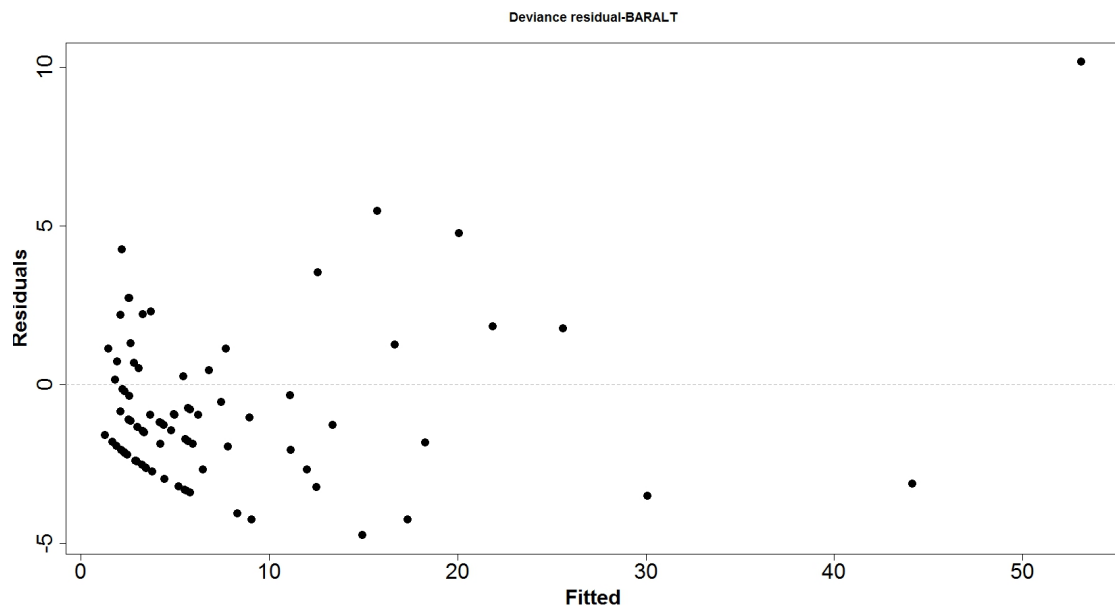


Figure C.17: Deviance of GAMM with random effect - Baralt

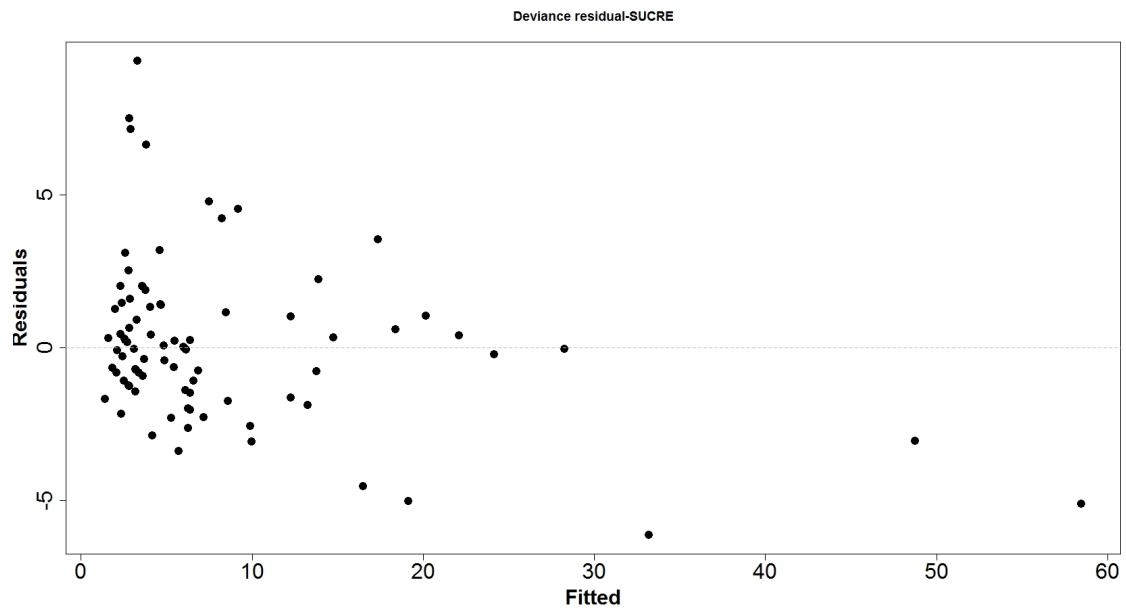


Figure C.18: Deviance of GAMM with random effect - Sucre

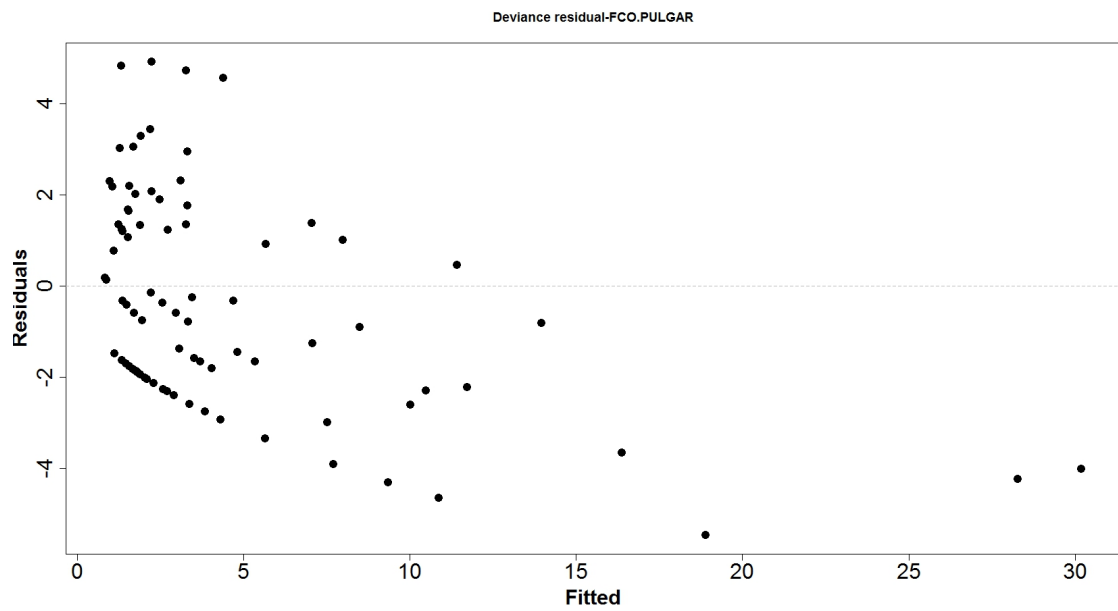


Figure C.19: Deviance of GAMM with random effect - Pulgar

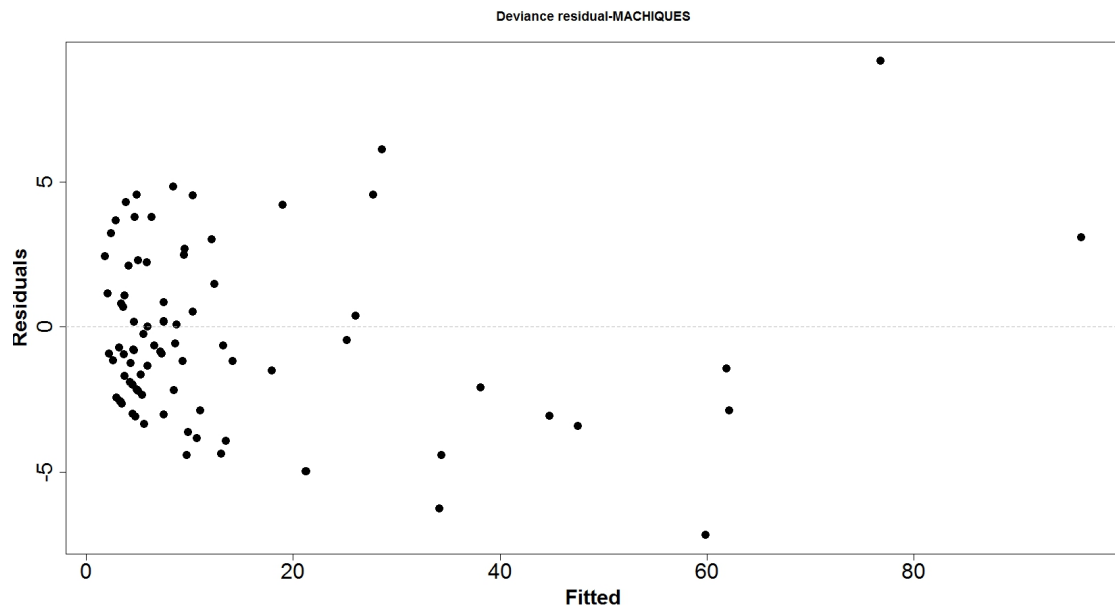


Figure C.20: Deviance of GAMM with random effect - Machiques

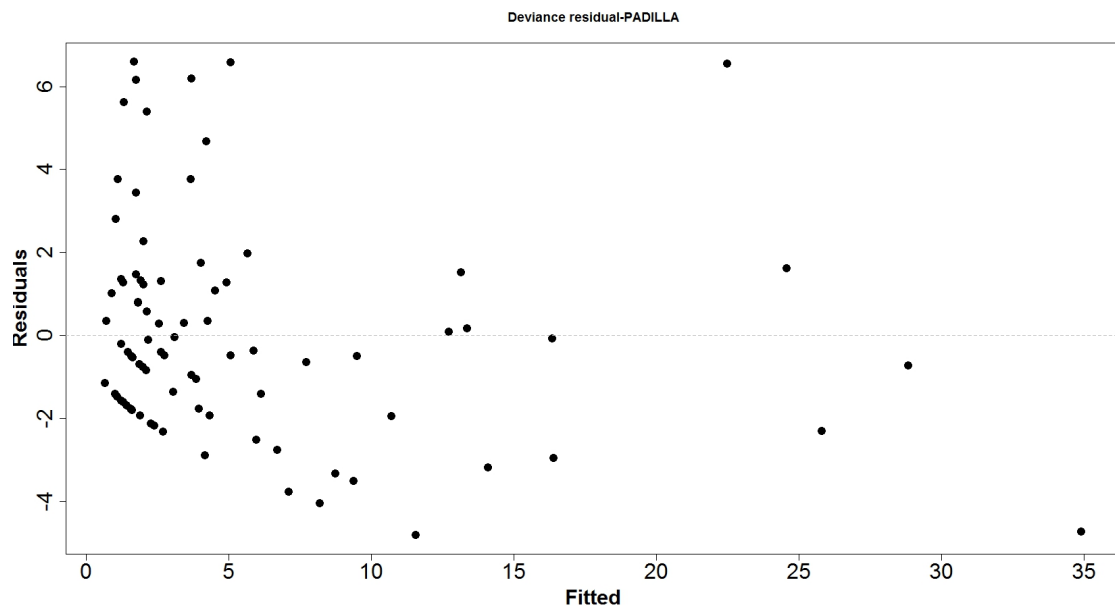


Figure C.21: Deviance of GAMM with random effect - Padilla

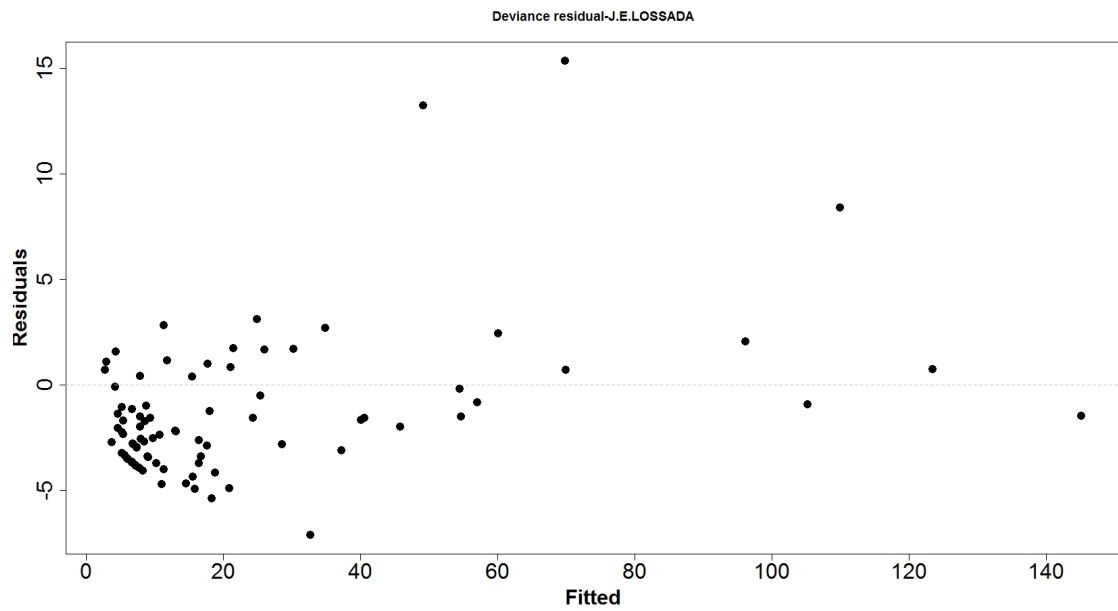


Figure C.22: Deviance of GAMM with random effect - Lossada

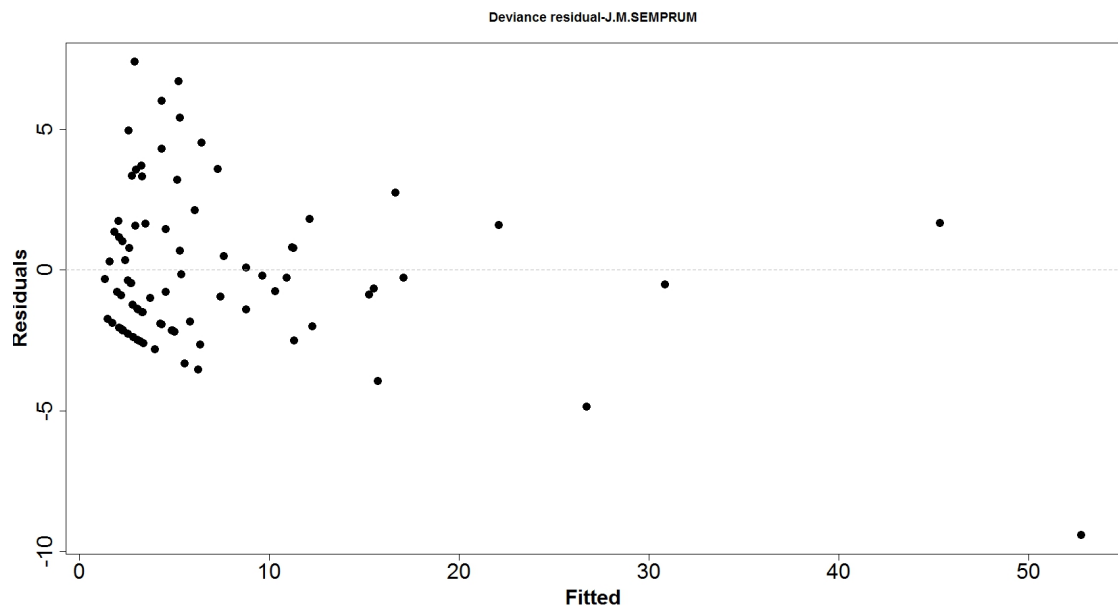


Figure C.23: Deviance of GAMM with random effect - Semprum

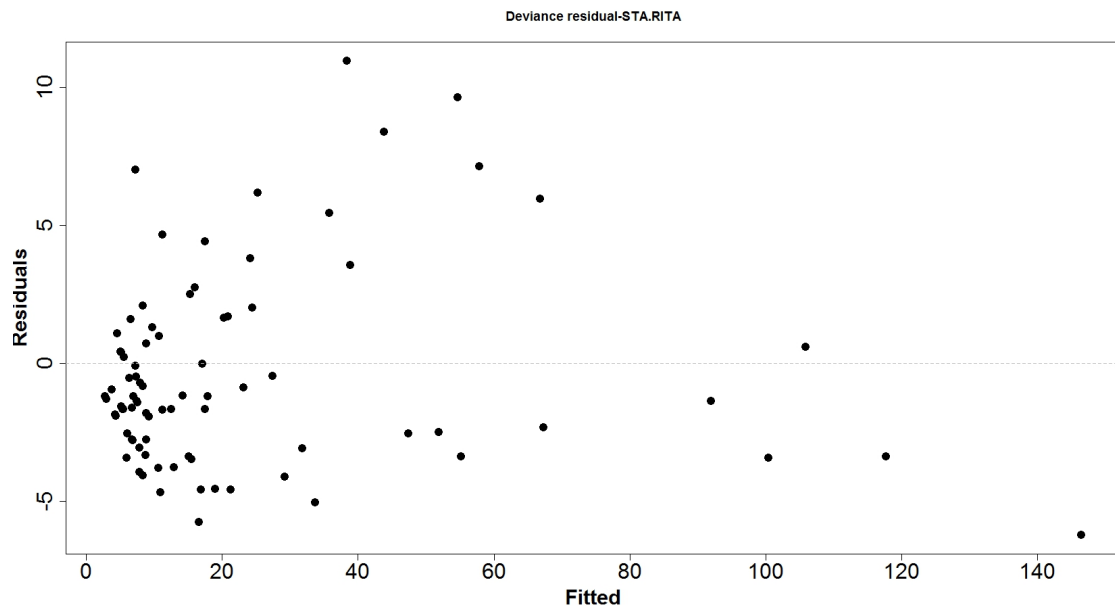


Figure C.24: Deviance of GAMM with random effect - Rita

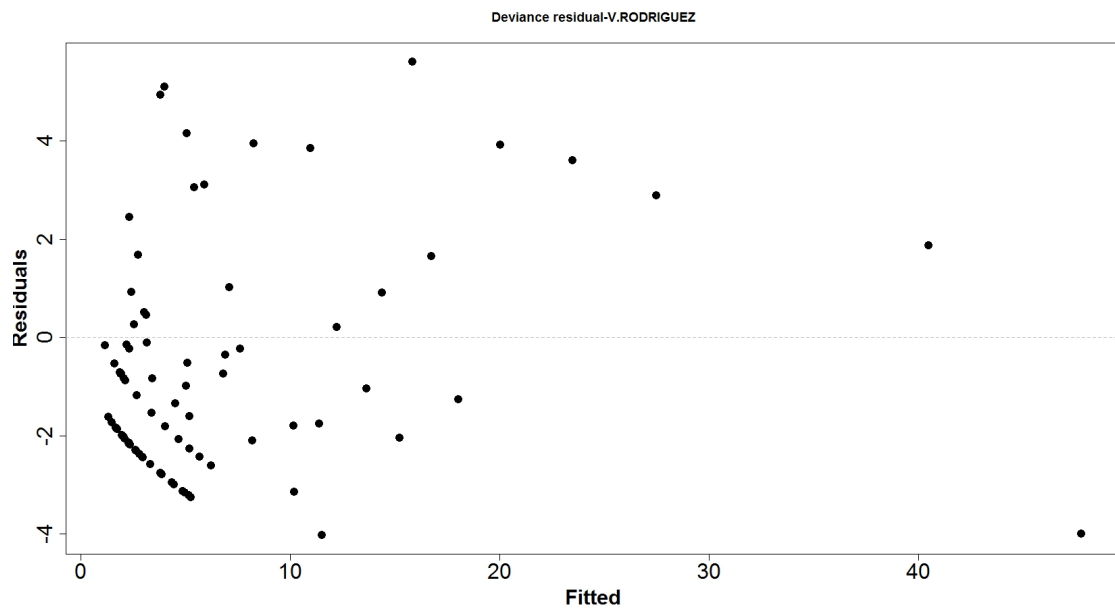


Figure C.25: Deviance of GAMM with random effect - Rodriguez

The following R-code shows the Moran' I test to assess for spatial autocorrelation. This code also allows a geographical representation of the residuals over 2002 to 2008 at municipality level in Zulia state.

```
#####
#####      SPATIAL DATA ANALYSIS      #####
#####      ZULIA STATE      #####
#####      Contiguity neighbours      #####
#####      by M. Cabrera      #####
#####

library (spdep)      # allow for spatial dependence
library(maptools)      # loads sp library tool
library(RColorBrewer) # creates nice color schemes
library(classInt)      # finds class intervals for continuous variables
#####

####      Zulia data (polygons)      #####
#####

municipios_limprios_zulia.shp <- readShapePoly(file.choose(),
proj4string=CRS("+proj=longlat"))
newdat<-read.csv("H:/dos/morantest2.csv",header=TRUE)
#####

##      Construction of a (nxn) spatial proximity matrix
##      CASEI: Neighbours based on Queen contiguity
#####

## Queen contiguity: polygons a common border or corner
#####

sc.01nb<-poly2nb(municipios_limprios_zulia.shp,queen=TRUE)
sc.01nb
summary(sc.01nb)
#####

### Bind data from newdat to municipios_limprios_zulia.shp ###
#####

##### Builds a neighbours based on Queen contiguity
##### Spatial weights for neighbours lists
sc.02nb<-nb2listw(sc.01nb,style="W")
#####      Residual from the GAM with random effect #####
```

```
##### RESIDUAL- 2002
  moran.test(newdat$gammresd2002,listw=sc.02nb)
##### RESIDUAL- 2003
  moran.test(newdat$gammresd2003,listw=sc.02nb)
##### RESIDUAL- 2004
  moran.test(newdat$gammresd2004,listw=sc.02nb)
##### RESIDUAL- 2005
  moran.test(newdat$gammresd2005,listw=sc.02nb)
##### RESIDUAL- 2006
  moran.test(newdat$gammresd2006,listw=sc.02nb)
##### RESIDUAL- 2007
  moran.test(newdat$gammresd2007,listw=sc.02nb)
##### RESIDUAL- 2008
  moran.test(newdat$gammresd2008,listw=sc.02nb)
```

```
#####
####   Display of Residuals per year- Visualizing the presence   ####
####   of spatial autocorrelation                               #####
#####
#####   RESIDUAL 2002
#####   SELECT OWN BREAKS

newdat<-read.csv("H:/dos/morantest2.csv",header=TRUE)
plotvar <-newdat$gammresd2002

plotclr <- brewer.pal(6,"OrRd")
class <- classIntervals(plotvar, style="fixed",fixedBreaks=c
  (-0.19,0,1.19,5.19,8.19,10.19,14.89))
colcode <- findColours(class, plotclr)
plot(municipios_limprios_zulia.shp, xlim=c(-73.25, -70), ylim=c(8,12))
plot(municipios_limprios_zulia.shp, col=colcode, add=T)
title(main="Normalized residual 2002",
  sub="GAMM with temporal autocorrelation structure")
legend(-70.10, 10.50, legend=names(attr(colcode, "table")),
  fill=attr(colcode, "palette"), cex=1.05, bty="n")
```

```
#####
#####   RESIDUAL 2003
#####   SELECT OWN BREAKS
#####

#####   SELECT OWN BREAKS

newdat<-read.csv("H:/dos/morantest2.csv",header=TRUE)
plotvar <-newdat$gammresd2003

plotclr <- brewer.pal(6,"OrRd")
class <- classIntervals(plotvar, style="fixed",fixedBreaks=c
  (-0.19,0,1.19,5.19,8.19,10.19,14.89))
colcode <- findColours(class, plotclr)
```

```
plot(municipios_limpios_zulia.shp, xlim=c(-73.25, -70), ylim=c(8,12))
plot(municipios_limpios_zulia.shp, col=colcode, add=T)
title(main="Normalized residual 2003",
      sub="GAMM with temporal autocorrelation structure")
legend(-70.10, 10.50, legend=names(attr(colcode, "table")),
      fill=attr(colcode, "palette"), cex=1.05, bty="n")
```

```
#####
##### RESIDUAL 2004
##### SELECT OWN BREAKS
#####

##### SELECT OWN BREAKS

newdat<-read.csv("H:/dos/morantest2.csv",header=TRUE)
plotvar <-newdat$gammresd2004
plotclr <- brewer.pal(6,"OrRd")
class <- classIntervals(plotvar, style="fixed",fixedBreaks=c
  (-0.19,0,1.19,5.19,8.19,10.19,14.89))
colcode <- findColours(class, plotclr)
plot(municipios_limprios_zulia.shp, xlim=c(-73.25, -70), ylim=c(8,12))
plot(municipios_limprios_zulia.shp, col=colcode, add=T)
title(main="Normalized residual 2004",
  sub="GAMM with temporal autocorrelation structure")
legend(-70.10, 10.50, legend=names(attr(colcode, "table")),
  fill=attr(colcode, "palette"), cex=1.05, bty="n")
```

```
#####
##### RESIDUAL 2005
##### SELECT OWN BREAKS
#####

newdat<-read.csv("H:/dos/morantest2.csv",header=TRUE)
plotvar <-newdat$gammresd2005
plotclr <- brewer.pal(6,"OrRd")
class <- classIntervals(plotvar, style="fixed",fixedBreaks=c
  (-0.19,0,1.19,5.19,8.19,10.19,14.89))
colcode <- findColours(class, plotclr)
plot(municipios_limprios_zulia.shp, xlim=c(-73.25, -70), ylim=c(8,12))
plot(municipios_limprios_zulia.shp, col=colcode, add=T)
title(main="Normalized residual 2005",
  sub="GAMM with temporal autocorrelation structure")
```

```
legend(-70.10, 10.50, legend=names(attr(colcode, "table")),  
fill=attr(colcode, "palette"), cex=1.05, bty="n")
```



```
#####
#####   RESIDUAL 2006
#####   SELECT OWN BREAKS
#####

newdat<-read.csv("H:/dos/morantest2.csv",header=TRUE)
plotvar <-newdat$gammresd2006
plotclr <- brewer.pal(6,"OrRd")
class <- classIntervals(plotvar, style="fixed",fixedBreaks=c
  (-0.19,0,1.19,5.19,8.19,10.19,14.89))
colcode <- findColours(class, plotclr)
plot(municipios_limprios_zulia.shp, xlim=c(-73.25, -70), ylim=c(8,12))
plot(municipios_limprios_zulia.shp, col=colcode, add=T)
title(main="Normalized residual 2006",
  sub="GAMM with temporal autocorrelation structure")
legend(-70.10, 10.50, legend=names(attr(colcode, "table")),
  fill=attr(colcode, "palette"), cex=1.05, bty="n")
```

```
#####
#####   RESIDUAL 2007
#####   SELECT OWN BREAKS
#####

newdat<-read.csv("H:/dos/morantest2.csv",header=TRUE)
plotvar <-newdat$gammresd2007
plotclr <- brewer.pal(6,"OrRd")
class <- classIntervals(plotvar, style="fixed",fixedBreaks=c
  (-0.19,0,1.19,5.19,8.19,10.19,14.89))
colcode <- findColours(class, plotclr)
plot(municipios_limprios_zulia.shp, xlim=c(-73.25, -70), ylim=c(8,12))
plot(municipios_limprios_zulia.shp, col=colcode, add=T)
title(main="Normalized residual 2007",
  sub="GAMM with temporal autocorrelation structure")
```

```
legend(-70.10, 10.50, legend=names(attr(colcode, "table")),  
      fill=attr(colcode, "palette"), cex=1.05, bty="n")
```

```
#####
##### RESIDUAL 2008
##### SELECT OWN BREAKS
#####

newdat<-read.csv("H:/dos/morantest2.csv",header=TRUE)
plotvar <-newdat$gammresd2008
plotclr <- brewer.pal(6,"OrRd")
class <- classIntervals(plotvar, style="fixed",fixedBreaks=c
  (-0.19,0,1.19,5.19,8.19,10.19,14.89))
colcode <- findColours(class, plotclr)
plot(municipios_limprios_zulia.shp, xlim=c(-73.25, -70), ylim=c(8,12))
plot(municipios_limprios_zulia.shp, col=colcode, add=T)
title(main="Normalized residual 2008",
  sub="GAMM with temporal autocorrelation structure")
legend(-70.10, 10.50, legend=names(attr(colcode, "table")),
  fill=attr(colcode, "palette"), cex=1.05, bty="n")
```

C.3 Third stage: GAMM for predictive purposes

Model specification

The R-codes specification of the three options proposed for predictive purposes are listed as follow:

```
library(mgcv)
library(MASS)

#####
##### First option: using interaction term between mean rainfall and
##### maximum temperature
#####

newdat$municipality<-as.factor(newdat$municipality)
newdat$saltgrp <-as.factor(newdat$saltgrp)
newdat$perhwatergrp <-as.factor(newdat$perhwatergrp)
```

```

First option<-gam(obsdengue~s(municipality,bs="re")+
                    s(month,bs="cc")+
                    altgrp+perhwatergrp+
                    s(meanrainfratelaglmm,bs="cs",k=3)+
                    s(nino3.4lag6,bs="cs",k=3)+
                    s(maxtempclag1,bs="cs",k=3)+
                    te(meanrainfratelaglmm,maxtempclag1,k=3),
                    offset=log(popsize), family=poisson,
                    data=newdat,scale=-1,gamma=1.4)

```

```
#####
##### Second option: using main effects on the climatic factors
#####
```

```
Second option<-gam(obsdengue~s(municipality,bs="re")+
                    s(month,bs="cc")+
                    altgrp+perhwatergrp+
                    s(meanrainfratelag1mm,bs="cs",k=3)+
                    s(nino3.4lag6,bs="cs",k=3)+
                    s(maxtempclag1,bs="cs",k=3),
                    offset=log(popsize), family=poisson,
                    data=newdat,scale=-1,gamma=1.4)
```

```
#####
##### Third option:
##### interaction term between mean rainfall and maximum temperature
##### and the El Nino3.4 in a categorical form
#####
```

```
Third option<-gam(obsdengue~s(municipality,bs="re")+
                    s(month,bs="cc")+
                    altgrp+perhwatergrp+groupnino+
                    s(meanrainfratelag1mm,bs="cs",k=3)+
                    te(meanrainfratelag1mm,maxtempclag1,k=3)+
                    s(maxtempclag1,bs="cs",k=3),
                    offset=log(popsize), family=poisson,
                    data=newdat,scale=-1,gamma=1.4)
```

Bibliography

- Abeku, T., Hay, S., Ochola, S., Langi, P., Bears, B., de Vlas, S., and Cox, J. (2004). Malaria epidemic early warning and detection on African highlands. *TRENDS in Parasitology*, 20(9):400–405.
- Abellan, J. and Best, R. (2008). Use of space-time models to investigate the stability of patterns of disease. *Environ Health Perspect*, 116(8):1111–1119.
- Acker, J. and Leptoukh, G. (2007). Online analysis enhances use of NASA earth science data. *EoS, Trans. AGU*, 88(2):14–17.
- Almeida, A. S. d., Medronho, R. d. A., and Valencia, L. I. O. (2009). Spatial analysis of dengue and the socioeconomic context of the city of Rio de Janeiro (Southeastern Brazil). *Rev Saude Publica*, 43(4):666–673.
- Alvarez, P. (2009). *X Plan de desarrollo del estado Zulia*. Gobernacion del estado Zulia.
- Anez, G., Balza, R., Valero, N., and Larreal, Y. (2006). Impacto economico del dengue y del dengue hemorragico en el estado Zulia, Venezuela, 1997 a 2003. *Rev Panam Salud Publica*, 19(5):314–320.
- Arboleda, S. and Peterson, A. T. (2009). Mapping environmental dimensions of dengue fever transmission risk in the Aburra Valley, Colombia. *Int. J. Environ. Res. Public Health*, 6(12):3040–3055.
- Ban, K. (2007). The Millennium Development goals report. Technical report, United Nations, New York.

- Banerjee, S., Carlin, B., and Gelfand, A. (2004). *Hierarchical modeling and analysis for spatial data*. Chapman & Hall CRC, Boca Raton, U.S.
- Barrera, R., Delgado, N., Jimnez, M., and Valero, S. (2002). Eco-epidemiological factors associated with hyperendemic dengue haemorrhagic fever in Maracay city, Venezuela. *Dengue Bulletin*, 26:84–95.
- Barrera, R., Delgado, N., Jimnez, M., Villalobos, I., and Romero, I. (2000). Estratificación de una ciudad hiperendémica en dengue hemorrágico. *Pan Am J Public Health*, 8:225–233.
- Barreto, F. R., Teixeira, M. G., Costa, M. C., Carvalho, M. S., and Barreto, M. L. (2008). Spread pattern of the first dengue epidemic in the city of Salvador, Brazil. *BMC Public Health*, 8(1).
- Basanez, M., Marshall, C., Carabin, H., Gyorkos, T., and Joseph, L. (2005). Bayesian statistics for parasitologists. *TRENDS Parasitol*, 20(2):85–91.
- Beck, L., Lobitz, B., and Wood, B. (2000). Remote sensing and human health: New sensors and new opportunities. *Emerging Infectious Diseases*, 6(3).
- Bennett, S., Drummond, A., Kapan, D., Suchard, M., Munoz, J., Pybus, O., Holmes, E., and Gubler, D. (2009). Epidemic dynamics revealed in dengue evolution. *Mol.Biol.Evol.*, 27(4):811–818.
- Berg, D. (2005). Bankruptcy prediction by Generalized Additive Models. Statistical Research Report 1, Dept. of Math. University of Oslo.
- Bivand, R. (2007). Analysing spatial data in R: Worked example: spatial autocorrelation.
- Bohra, A. and Andrianasolo, H. (2001). Application of GIS in modeling of dengue risk based on sociocultural data: case of Jalore, Rajasthan, India. *Dengue Bull*, 25:92–102.

- Boscoe, F., Ward, M., and Reynolds, P. (2004). Current practices in spatial analysis of cancer data: data characteristics and data sources for geographic studies of cancer. *Int J Health Geogr*, 3(28).
- Bouma, M., Poveda, G., Rojas, W., Chavasse, D., Quiones, M., Cox, J., and J., P. (1997). Predicting high-risk years for malaria in Colombia using parameters of El Nino Southern Oscillation. *Trop Med Int Health*, 2(12):1122–1127.
- Bouma, M. and Van der Kaay, H. (1996). The El Nino Southern Oscillation and the historic malaria epidemics on the Indian subcontinent and Sri Lanka: an early warning system for future epidemics? *Trop Med Int Health*, 1(1):86–96.
- Brightmer, M. and Fantato, M. (1998). Human and environmental factors in the increasing incidence of dengue fever: a case study from Venezuela. *GeoJournal*, 44(2):103–109.
- Brunkard, J. M., Cifuentes, E., and Rothenberg, S. J. (2008). Assessing the roles of temperature, precipitation, and ENSO in dengue re-emergence on the Texas Mexico border region. *Salud Publica Mex.*, 50(3):227–234.
- Cameron, A. and Trivedi, P. (1999). Essentials of count data regression. Technical report.
- Caprara, A., Lima, J. W. d. O., Marinho, A. C. P., Calvasina, P. G., Landim, L. P., and Sommerfeld, J. (2009). Irregular water supply, household usage and dengue: a bio-social study in the Brazilian Northeast. *Cad Sade Pblica*, 25:S125–S136.
- Carbajo, A., Cardo, M., and Vezzani, D. (2012). Is temperature the main cause of dengue rise in non-endemic countries? the case of Argentina. *International Journal of Health Geographics*.
- Ceccato, P., Ghebremeskel, T., Jaiteh, M., Graves, P. M., Levy, M., Ghebreselassie, S., Ogbamariam, A., Barnston, A. G., Bell, M., and del Corral, J. (2007). Malaria stratification, climate, and epidemic early warning in Eritrea. *Am J Trop Med Hyg*, 77(6):61–68.

- Chan, M. (2012). Global strategy for dengue prevention and control. Technical Report WC-528, WHO, Geneva, Switzerland.
- Chang, A. Y., Parrales, M. E., Jimenez, J., Sobieszczyk, M. E., Hammer, S. M., Copenhagen, D. J., and Kulkarni, R. P. (2009). Combining Google Earth and GIS mapping technologies in a dengue surveillance system for developing countries. *International Journal of Health Geographics*, 8(1):49.
- Chen, C., Liao, C., Chio, C., Chou, H., You, S., and Cheng, Y. (2010). Lagged temperature effect with mosquito transmission potential explains dengue variability in southern Taiwan: Insights from a statistical analysis. *Sci Total Environ*, 408(19):4069–4075.
- Cheung, Y. (2002). Zero-inflated models for regression analysis of count data: a study of growth and development. *Statist. Med*, 21.
- Chi, Y., Zhonguei, Y., and Yueheng, S. (2012). Probabilistic precipitation forecasting based on ensemble output using generalized additive models and Bayesian model averaging. *Acta Meteor. Sinica*, 26(1).
- Chien, C. and Chen, T. (2009). A Bayesian model for age, period, and cohort effects on mortality trends for lung cancer, in association with gender-specific incidence and case fatality rates. *J Thorac Oncol*, 4(2).
- Cleophas, T., Zwinderman, A., Atiqi, R., and Bosh, J. (2010). Cardiovascular research: data dispersion issues. Technical report, European Inter-university College Pharmaceutical Medicine.
- Clifford, S., Choy, S., Hussein, T., Mengersen, K., and Morawska, L. (2011). Using generalized additive model to model the particle number count of ultrafine particles. *Atmospheric Environment*, 45:5934–5945.
- Connolly, M. (2005). Communicable disease control in emergencies. a field manual. Technical report, WHO, France.

- Connor, S. J., Thomson, M. C., Flasse, S. P., and Perryman, A. H. (1998). Environmental information systems in malaria risk mapping and epidemic forecasting. *Disasters*, 22(1):39–56.
- Cordeiro, R., Donalisio, M., Andrade, V., Mafra, A., Nucci, L., Brown, J., and Stephan, C. (2011). Spatial distribution of the risk of dengue fever in southeast Brazil, 2006–2007. *BMC Public Health*.
- Das, A., Lele, S., Glass, G., Shields, G., and Patz, J. (2002). Modelling a discrete spatial response using generalized linear mixed models: application to Lyme disease vectors. *Int. J. Geographical Information Science*, 16(2):151–166.
- Descoux, E., Mangeas, M., Menkes, C., Lengaigne, M., Leroy, A., Tehei, T., Guillaumot, L., Teurlai, M., Gourinat, A., Benzler, J., Pfannstiel, A., Grangeon, J., Degallier, N., and Lamballerie, X. (2012). Climate based models for understanding and forecasting dengue epidemics. *Plos Negl Trop Dis*, 6(2).
- Diaz, M., Rodriguez, H., Rothenberg, S., Dantes, H., and Cifuentes, E. (2007). Impact of climatic variability on the incidence of dengue in Mexico. *Trop. Med and Int. Health*, 12(11):1327–1337.
- Dickin, S., Schuster, C., and Elliott, S. (2013). Developing a vulnerability mapping methodology: applying the water associated disease index to dengue in Malaysia. *PLOS ONE*, 8(5).
- Eisen, L. and Lozano-Fuentes, S. (2009). Use of mapping and spatial and space-time modeling approaches in operational control of *Aedes aegypti* and dengue. *PLoS neglected Tropical diseases*, 3(4).
- Elliott, P., Wakefield, J., Best, N., and Briggs, D. (2000). *Spatial Epidemiology Methods and Applications*. Oxford University Press, Oxford.
- Elliott, P. and Wartenberg, D. (2004). Spatial epidemiology: current approaches and future challenges. *Environ Health Perspect*, 112(9):998–1006.

- Epstein, P. R., Diaz, H. F., Elias, S., Grabherr, G., Graham, N. E., Martens, W. J., Mosley-Thompson, E., and Susskind, J. (1998). Biological and physical signs of climate change: focus on mosquito-borne diseases. *Bulletin of the American Meteorological Society*, 79(3):409–417.
- Ferreira, G. S. and Schmidt, A. M. (2006). Spatial modelling of the relative risk of dengue fever in Rio de Janeiro for the epidemic period between 2001 and 2002. *Brazilian J Probability Statistics*, 20:29–47.
- Freitas, M., Souris, P., Sibajev, A., Weimann, E., Marques, A., Ferreira, R., and Luitgards, J. (2003). Exploratory temporal and spatial distribution analysis of dengue notifications in Boa Vista, Roraima Brazilian Amazon, 1999-2001. *Dengue Bull*, 27:63–80.
- Fuller, D., Troyo, A., and Beier, J. (2009). El Nino Southern Oscillation and vegetation dynamics as predictors of dengue fever cases in Costa Rica. *Environ Res Lett.*, 4:140111–140118.
- Gagnon, A. S., Bush, A. B., and Smoyer-Tomic, K. E. (2001). Dengue epidemics and the El Nino Southern Oscillation. *Clim Res*, 19(1):35–43.
- Garelli, F., Espinosa, M., Weinberg, H., Coto, H., Gaspe, M., and Gurtler, R. (2009). Patterns of *Aedes aegypti*. infection and container productivity measured using pupal and *Stegomyia* indices in Northern Argentina. *Entomological Society of America*, 46(5):1176–1186.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1:1–19.
- Gelman, A. and Hill, J. (2007). *Data analysis using regression and Multilevel Hierarchical*. Cambridge University Press.
- Gibbons, R. V. and Vaughn, D. W. (2002). Dengue: an escalating problem. *BMJ: British Medical Journal*, 324(7353):1563–1566.

- Jimenez, O., Bonner, S., Bonner, J., King, R., Parker, R., Brooks, S., Jamieson, L., Grosbois, V., Morgan, B., and Thomas, L. (2009). WinBUGS for population ecologists bayesian modeling using Markov Chain Monte Carlo methods. *Environmental and Ecological statistics*, pages 883–915.
- GIS, Q. (2009). *Quantum GIS User, Installation and Coding Guide*, 1.3 edition.
- Githeko, A. K., Lindsay, S. W., Confalonieri, U. E., and Patz, J. A. (2000). Climate change and vector-borne diseases: a regional analysis. *Bulletin of the World Health Organization*, 78(9):1136–1147.
- Greenland, S. (2006). Bayesian perspective for epidemiological research: I foundations and basic methods. *International Journal of Epidemiology*, 35:765–775.
- Gschlobl, S. and Czado, C. (2008). Modelling count data with overdispersion and spatial effects. *Statistical papers*, 49:531–552.
- Gubler, D. J. (1998a). Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev*, 11(3):480–496.
- Gubler, D. J. (1998b). Resurgent vector-borne diseases as a global health problem. *Emerging infectious diseases*, 4(3):1–11.
- Guha, S. and Ryan, L. (2006). Spatio-temporal analysis of areal data and discovery of neighborhood relationships in Conditionally Autoregressive Models. Technical report, Harvard University, Harvard University Biostatistics working Paper.
- Guha-Sapir, D. and Schimmer, B. (2005). Dengue fever: new paradigms for a changing epidemiology. *Emerg Themes Epidemiol*, 2(1).
- Guisan, A., Edwards, T., and Hastie, T. (2002). Generalized linear and generalized additive models in studies of species distributions: setting the scene. *Ecological Modelling*, 157:89–100.
- Gurtler, R., Garelli, F., and Coto, H. (2009). Effects of a five-year citywide intervention program to control aedes aegypti and prevent dengue outbreaks on Northern Argentina. *PLoS Neglected Tropical Diseases*, 3.

- Hales, S., Weinstein, P., Souares, Y., and Woodward, A. (1999). El Nino and the dynamics of vectorborne disease transmission. *Environmental Health Perspectives*, 107(2):99–102.
- Halsey, E., Marks, M., Gottuzo, E., Fiestas, V., Suarez, L., Vargas, J., Aguayo, N., Madrid, C., Vimos, C., Kochel, T., and Laguna, A. (2012). Correlation of serotype-specific dengue virus infection with clinical manifestations. *PLoS Neglected Tropical diseases*, 6(5).
- Hara, R. (2009). How to make models add up- a primer on GLMMs. *Ann. Zool. Fennici*, 46:124–137.
- Hastie, T. and Tibshirani, R. (1984). Generalized additive models. Technical report 2, Laboratory for Computational Statistics, Stanford University.
- Hastie, T. and Tibshirani, R. (1986). Generalized additive models. *Statistical Science*, 1(3):297–318.
- Henchal, E. A. and Putnak, J. R. (1990). The dengue viruses. *Clinical microbiology reviews*, 3(4):376–396.
- Hii, Y., Zhu, H., Ng, N., Ng, L., and Rocklov, J. (2012). Forecast of dengue incidence using temperature and rainfall. *PLoS Negl Trop Dis*, 6(11).
- Hii, Y. L., Joacim, R., Nawi, N., Choon Siang, Tang, F. Y. P., and Rainer, S. (2009). Climate variability and increase in intensity and magnitude of dengue incidence in Singapore. *Glob Health Action*, 2(10):100–200.
- Hoeff, J. and Boveng, P. (2007). Quasi Poisson vs. Negative Binomial regression. How we should model overdispersed data. *Ecology*, 88(11):2766–2772.
- Honorio, N., Nogueira, R., Torres, C., Carvalho, M., Goncalvez, O., Magalhaes, M., Araujo, J., Araujo, E., Gomes, M., Pinheiro, I., Pinel, C., and Oliveira, R. (2009). Spatial evaluation and modelling of dengue seroprevalence and vector density in Rio de Janeiro, Brazil. *PLoS Negl Trop Dis*, 3(11).

- Hopp, M. J. and Foley, J. A. (2001). Global-scale relationships between climate and the dengue fever vector, *Aedes aegypti*. *Climatic Change*, 48:441–463.
- Hu, W., Clements, A., Williams, G., and Tong, S. (2010). Dengue fever and El Niño/Southern Oscillation in Queensland, Australia: a time series predictive model. *Occup Environ Med*, 67(5):307–311.
- Ismaila, A., Canty, A., and Thabane, L. (2007). Comparison of Bayesian and frequentist approaches in modelling risk of preterm birth near the Sydney Tar Ponds, Nova Scotia, Canada. *BMC Medical Research Methodology*, 7(39).
- Jarup, L. (2004). Health and environment information systems for exposure and disease mapping, and risk assessment. *Environ Health Perspect*, 112(9):995–9977.
- Johansson, M. A., Dominici, F., and Glass, G. E. (2009). Local and global effects of climate on dengue transmission in Puerto Rico. *PLoS Negl Trop Dis*, 3(2).
- Kempler, S. (2013). Random effects in gams. http://gdata1.sci.gsfc.nasa.gov/daac-bin/G3/gui.cgi?instance_id=mairs_monthly, Accessed: 26 June 2012.
- Kery, M. (2010). *Introduction to WinBUGS for Ecologists: Bayesian approach to regression, ANOVA, mixed models and related analyses*. Academic Press, San Diego, USA.
- Kim, H. and Lim, H. (2010). Comparison of Bayesian Spatio-Temporal models for chronic diseases. *Journal of Data Science*, 8:189–211.
- Kongsomboon, K., Singhasivanon, P., Kaewkungwal, J., Nimmannitya, S., Mammen, M., Nisalak, A., and Sawanpanyalert, P. (2004). Temporal trends of dengue fever/-dengue hemorrhagic fever in Bangkok, Thailand from 1981 to 2000: an age-period-cohort analysis. *Southeast Asian J Trop Med Public Health*, 35(4):913–917.
- Kovats, R. S. (2000). El Niño and human health. *Bull World Health Organ*, 78(9):1127–35.
- Kovats, R. S., Bouma, M. J., Hajat, S., Worrall, E., and Haines, A. (2003). El Niño and health. *The Lancet*, 362(9394):1481–1489.

- Kyle, J. L. and Harris, E. (2008). Global spread and persistence of dengue. *Annu. Rev. Microbiol.*, 62:71–92.
- Lado, M., Cadarso-Surez, C., Roca-Pardias, J., and Tahoces, P. (2008). Categorical variables, interactions and generalized additive models. Applications in computer-aided diagnosis systems. *Comput Biol Med*, 38:475–483.
- Lawson, A. (2009). *Bayesian Disease mapping, Hierarchical modelling in Spatial epidemiology*. Chapman & Hall CRC, Boca Raton, U.S.
- Lawson, A., Biggeri, A., Boehning, D., Lesaffre, E., Viel, J., Clark, A., Schlattmann, P., and Divino, F. (2000). Disease mapping models: an empirical evaluation. *Statist. Med.*, 19:2217–2241.
- Lawson, A., Browne, W., and Rodeiro, C. (2003). *Disease mapping with WinBUGS and MLwiN*. Wiley John & Sons, England.
- Lawson, A., Clark, A., and Vidal, C. (2004). Developments in general and Syndromic Surveillance for small area health data. *Journal of Applied Statistics*, 31:951–966.
- Lawson, A. and Fiona, W. (2001). *An Introduction guide to Disease Mapping*. Wiley & Sons, Ltd, West Sussex, England.
- Lee, D. and Mitchell, R. (2012). Locally adaptive spatial smoothing using conditional autoregressive models. Technical report, School of Mathematics and Statistics, University of Glasgow, Glasgow.
- Lovett, A. and Flowerdew, R. (1989). Analysis of count data using Poisson regression. *Professional Geographer*, 41(2):190–198.
- Lowe, R., Bailey, T., Stephenson, D., Jupp, T., Graham, R., Barcellos, C., and Carvalho, M. (2013). The development of an early warning system for climate-sensitive disease risk with a focus on dengue epidemics in Southeast Brazil. *Statist. Med*, 32:864883.

- Lowe, R., Bailey, T. C., Stephenson, D. B., Graham, R. J., Coelho, C., Carvalho, M., and Barcellos, C. (2011). Spatio-temporal modelling of climate-sensitive disease risk: Towards an early warning system for dengue in Brazil. *Computers & Geosciences*, 37(3):371–381.
- Lozano, S., Elizondo-Quiroga, D., Farfan, J. A., Lorono-Pino, M. A., Garcia, J., Gomez-Carro, S., Lira, V., Najera-Vazquez, R., Fernandez, I., and Calderon-Martinez, J. (2008). Use of Google EarthTM to strengthen public health capacity and facilitate management of vector-borne diseases in resource-poor environments. *Bull World Health Organ*, 86(9):718–725.
- Lunn, D., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS a Bayesian modelling framework: Concepts, structure and extensibility. *Statistics and Computing*, 10:325–337.
- Luo, Y. (2012). *Novel approaches and their applications in risk assessment*. InTech.
- Mammen, M. P., Pingate, C., Koenraad, C. J., Rothman, A. L., Aldstadt, J., Nisalak, A., Jarman, R. G., Jones, J. W., Srikiatkachorn, A., and Ypil-Butac, C. A. (2008). Spatial and temporal clustering of dengue virus transmission in Thai villages. *PLoS Medicine*, 5(11):1605–1616.
- Martinez, M., Lopez, A., and Botella, P. (2008). An autoregressive approach to spatio-temporal disease mapping. *Stat Med*, 27(15):2874–2889.
- McMahon, P., Zaslavsky, A., Weinstein, M., Kuntz, K., Weeks, J., and Gazelle, G. (2006). Estimation of mortality rates for disease simulation models using Bayesian evidence synthesis. *Med Decis Making*, 26(5):497–511.
- McPhaden, M. (2008). Evolution of the 2006-2007 El Nino: the role of intraseasonal to interannual time scale dynamics. *Adv Geosci*, 14:219–230.
- Ministerio (2007). Aspectos generales sobre el dengue. clinica, diagnostico, tratamiento y control.

- Mojica, J., Ferreira, J., Escalante, M., Barrientos, E., Diaz, S., Guiffrida, A., Velarde, A., Calderon, I., Garcia, M., Zambrano, J., and Cedeno, N. (2012). Incidencia y circulacion del virus dengue en el estado Zulia, Venezuela 2009-2010. *CIENCIA*, 20(1):22–32.
- Mondini, A. and Chiaravalloti-Neto, F. (2008). Spatial correlation of incidence of dengue with socioeconomic, demographic and environmental variables in a Brazilian city. *Science of the Total Environment*, 393(2-3):241–248.
- Monsalve, N., Rubio, Y., and Maria, P. (2010). Modelo bayesiano espacio temporal de factores asociados con la incidencia del dengue en el area metropolitana de Maracay, Venezuela. *Boletin de Malariologia y salud ambiental*, L(2):219–232.
- Morrison, A. C., Getis, A., Santiago, M., Rigau-Perez, J. G., and Reiter, P. (1998). Exploratory space-time analysis of reported dengue cases during an outbreak in Florida, Puerto Rico, 1991-1992. *American Journal of Tropical Medicine and Hygiene*, 58(3):287–298.
- Mullen, R. and Birkeland, K. (2008). Mixed effects and spatial correlation models for analyzing a regional spatial dataset. Technical report, Proceedings of the 2008 International Snow Science Workshop, British Columbia.
- Murase, H., Nagashima, H., Yonezaki, S., Matsukura, R., and Kitakado, T. (2009). Application of a generalized additive model (GAM) to reveal relationships between environmental factors and distributions of pelagic fish and krill: a case study in Sendai Bay, Japan. *ICES Journal of Marine Science*, 66:14171424.
- Muttitanon, W., Kongthong, P., Kongkanon, C., Yoksan, S., Nitatpattana, N., Gonzales, J., and Barbazan, P. (2004). Spatial and temporal dynamics of dengue hemorrhagic fever epidemics, Nakhon Pathom province, Thailand, 1997-2001. *Den Bull*, 28:35–43.
- Nakhapakorn, K. and Tripathi, N. (2005). An information value based analysis of physical and climatic factors affecting dengue fever and dengue haemorrhagic fever incidence. *Int J of Health Geogr*, 4(13).

- Ntzoufras, I. (2009). *Bayesian Modeling using WINBUGS, An Introduction*. John Wiley & Sons, Inc., New Jersey.
- Oletta, L. (2006). Dengue en America Latina y Venezuela. *Med Interna*, 22:247–258.
- Osei, P. (2010). *Statistical Methods for Disease Mapping*. PhD thesis, African Institute for Mathematical Sciences.
- Ostfeld, R. S., Glass, G. E., and Keesing, F. (2005). Spatial epidemiology: an emerging (or re-emerging) discipline. *Trends in Ecology and Evolution*, 20(6):328–336.
- Patz, J. A., Martens, W., Focks, D. A., and Jetten, T. (1998). Dengue fever epidemic potential as projected by general circulation models of global climate change. *Environ Health Perspect*, 106(3):147–153.
- Pfeiffer, D. (1996). Issues related to handling of spatial data. pages 83–105.
- Pfeiffer, D., Robinson, T., Stevenson, M., Stevens, K., Rogers, D., and Clements, A. (2008). *Spatial Analysis in Epidemiology*. Oxford University Press, New York.
- Porcasi, X., Rotela, C., Introini, M., Frutos, N., Lanfri, S., Peralta, G., De Elia, E., Lanfri, M., and Scavuzzo, C. (2012). An operative dengue risk stratification system in Argentina based on geospatial technology. *Geospat Health*.
- Racloz, V., Rebecca, R., Tong, S., and Hu, W. (2012). Surveillance of Dengue fever virus: A review of Epidemiological models and Early Warning Systems. *PLoS Negl Trop Dis*, 6(5).
- Raja, N. and Devi, S. (2006). The incidence of dengue disease in a university teaching hospital in Malaysia in 2002, 2003 and 2004. *Infect Dis Pak*, 15:99–102.
- Rezaeian, M., Dunn, G., St Leger, S., and Appleby, L. (2007). Geographical epidemiology, spatial analysis and geographical information systems: a multidisciplinary glossary. *J Epidemiol Community Health*, 61(2):98–102.

- Richardson, S., Thomson, A., Best, N., and Elliott, P. (2004). Interpreting posterior relative risk estimates in disease mapping studies. *Environ Health Perspect*, 112(9):1016–1025.
- Rifakis, P., Goncalves, N., and Omaa, W. (2005). Asociacion entre las variaciones climaticas y los casos de dengue en un hospital de Caracas, Venezuela, 1998-2004. *Rev Per Med Exp Salud Publica*, 22(3):183–190.
- Riley, R., Higgins, J., and Deeks, J. (2011). Interpretation of random effects meta-analyses. *BMJ*, 342:964–967.
- Rodier, G., Chan, M., and Nafo, F. (2004). Using climate to predict infectious disease outbreaks: a review. Technical report, WHO, Geneva.
- Rodriguez, I., Sez Sez, V., Rubio-Palis, Y., and Vsquez, M. (2007). Estudio preliminar: zonas de amenaza epidemiologica de dengue bajo condiciones de estacionalidad de la lluvia (1997-2002). Area Metropolitana de Maracay, estado Aragua, Venezuela. *Terra. Nueva Etapa*, 23(33).
- Root, E. (2011). Time series analysis. stochastic processes. http://colorado.edu/geography/class_homepages/geog_4023_s11/Lecture16_TS3.pdf, Accessed: 02 January 2012.
- Rothman, K., Greenland, S., and Lash, T. (2008). *Modern Epidemiology*. Lippincott Williams & Wilkins, Philadelphia, USA.
- Saez, V. (2006). Estudio correlativo entre dengue, precipitacion y temperatura del aire, periodo 1995 a 2002. Municipio Libertador. Distrito Capital. Venezuela. *Terra.Nueva Etapa*, 22(32):123–156.
- SanMartin, J. L., Brathwaite, O., Zambrano, B., Solorzano, J. O., Bouckennooghe, A., Dayan, G. H., and Guzman, M. G. (2010). The epidemiology of dengue in the Americas over the last three decades: a worrisome reality. *Am J Trop Med Hyg*, 82(1):128–35.

- Sattenspiel, L. (2009). *The Geographic spread of infectious diseases Models and applications*. Princeton University Press, New Jersey, USA.
- Schukken, Y., Casella, G., and Broek, J. (1991). Overdispersion in clinical mastitis data from dairy herds: a negative binomial approach. *Preventive Veterinary Medicine*, 10:239–245.
- Schwartz, J. (1993). Air pollution and daily mortality in Birmingham, Alabama. *American Journal of Epidemiology*, 10:1136–1147.
- Sierra, B. d. l. C., Kouri, G., and Guzmán, M. (2007). Race: a risk factor for dengue hemorrhagic fever. *Arch Virol*, 152(3):533–542.
- Siqueira, J., Maciel, I., Barcellos, C., Souza, W., Carvalho, M., Nascimento, N., Oliveira, R., Morais, O., and Martelli, C. (2008). Spatial point analysis based on dengue surveys at household level in central Brazil. *BMC Public Health*.
- Solymosi, N., Wagner, S., Agots, A., and Allepuz, A. (2010). *maps2WinBUGS 2.2. QGIS plugin manual*.
- Soni, A., Chugh, K., Sachdev, A., and Gupta, D. (2001). Management of dengue fever in ICU. *Indian J Pediatr*, 68(11):1051–1055.
- Spiegelhalter, D. (1998). Bayesian graphical modelling: a case-study in monitoring health outcomes. *Appl. Statist.*, 47:115–133.
- Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). *WinBUGS User Manual Version 1.4*. Institute of Public Health Cambridge and Imperial College School of Medicine.
- Staelens, J., Wuyts, K., S., A., Van Avermaet, P., and H., B. (2012). Trends in atmospheric nitrogen and sulphur deposition in northern Belgium. *Atmospheric Environment*, 49:186–196.
- Stevenson, M. (2011). Investigation of Spatial Patterns of Animal Disease. Technical report, University Massey.

- Tan, S. (2001). Introduction to Bayesian methods for medical research. *Ann Acad Med Singapore*, 30:444–446.
- Terzi, Y. and Cengiz, M. (2009). Using of generalized additive model for model selection in multiple Poisson regression for air pollution data. *Sci. Res. and Essay*, 4(9):867–871.
- Thammapalo, S., Chongsuvivatwong, V., Geater, A., and Dueravee, M. (2008a). Environmental factors and incidence of dengue fever and dengue haemorrhagic fever in an urban area, Southern Thailand. *Epidemiology and infection*, 136(1):135–143.
- Thammapalo, S., Nagao, Y., Sakamoto, W., Saengtharatip, S., Tsujitani, M., Nakamura, Y., Coleman, P., and Davies, C. (2008b). Relationship between transmission intensity and incidence of dengue hemorrhagic fever in Thailand. *PLoS Negl Trop Dis*, 2(7).
- Thomas, A., Best, N., Lunn, D., Arnold, R., and Spiegelhalter, D. (2004). *GeoBUGS User Manual*.
- Thomas, H., Bailye, W., Amarakoon, D., Chen, A., Rawlins, S., Chadee, S., R, C., Owino, A., and Polson, K. (2006). Vulnerability to dengue fever in Jamaica. *AIACC Working paper*, (27).
- Thomson, M., Mason, S., Phindela, T., and Connor, S. (2004). Use of rainfall and sea surface temperature monitoring for malaria early warning in Botswana. *TRENDS in Parasitology*, 20(9):400–405.
- Trevisani, M. and Torelli, N. (2007). Hierarchical Bayesian models for small area estimation with count data. Technical report, Dipartimento di Scienze Economiche e Statistiche, Università Degli studi di Trieste, Trieste.
- Underwood, F. (2009). Describing long-term trends in precipitation using generalized additive models. *Journal of Hydrology*, 364:285–297.

- Vislocky, R. and Fritsch, M. (1995). Generalized additive models versus linear regression in generating probabilistic MOS forecasts of aviation weather parameters. *American Meteorological Society*, 10(4):669–680.
- Wakefield, J. (2004). A critique of statistical aspects of ecological studies in spatial epidemiology. *Environmental and Ecological Statistics*, 11:31–54.
- Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostat*, 8(2):158–183.
- Waller, L., Carlin, B., Xia, H., and Gelfand, A. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92(438):607–617.
- Wang, H. (2010). Closed form prediction intervals applied for disease counts. *The American Statistician*, 64(3).
- Wang, S. and Morishima, G. (2009). The use of generalized additive models for forecasting the abundance of Queets River Coho salmon. *North American Journal of fisheries management*, 29:423–433.
- Wayne, N. (1982). *Applied life data analysis*. John Wiley & Sons.
- Weiss, J. (2008). Predictive simulation.
- Weiss, J. (2010). Refit the frequentist and Bayesian analysis of covariance.
- Werneck, G. (2008). Georeferenced data in epidemiologic research. *Cienc Saude Coletiva*, 13:1753–1766.
- Wood, S. (2005). Low rank scale invariant tensor product smooths for Generalized Additive Mixed Models. Technical report, Department of Statistics, University of Glasgow, Glasgow.
- Wood, S. (2006). *Generalized additive models. An introduction with R*. Chapman & Hall CRC, Boca Raton, U.S.

- Wood, S. (2013). Random effects in gams. <http://stat.ethz.ch/R-manual/R-patched/library/mgcv/html/random.effects.html>, Accessed: 19 July 2012.
- Wu, P.-C., Lay, J.-G., Guo, H.-R., Lin, C.-Y., Lung, S.-C., and Su, H.-J. (2009). Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. *Science of the Total Environment*, 407(1).
- Yu, H.-L., Yang, S.-J., Yen, H.-J., and Christakos, G. (2011). A spatio-temporal climate-based model of early dengue fever warning in southern Taiwan. *Stoch Environ Res Risk Assess*, 25(4):485–494.
- Zacarias, O. and Andersson, M. (2011). Spatial and temporal patterns of malaria incidence in Monzambique. *Malaria Journal*, 10.
- Zeileis, A., Kleiber, C., and Jackman, S. (2008). Regression models for count data in R. *Journal of Statistical Software*, 27(8).
- Zuur, Ieno, Walker, Saveliev, and Smith (2009). *Mixed Effects Models and Extensions in Ecology with R*. Springer Science, New York.
- Zuur, A., Ieno, E., and Elphick, C. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology & Evolution*, 1:3–14.
- Zuur, A. F. (2012). *A Beginner's Guide to Generalized Additive Models with R*. Highland Statistics Ltd, Newburgh, UK.